79

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
31 May 2007 (31.05.2007)

PCT

(10) International Publication Number
WO 2007/062138 A2

(51) International Patent Classification:
G01N 33/53 (2006.01)

(21) International Application Number:
PCT/US2006/045237

(22) International Filing Date:
22 November 2006 (22.11.2006)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/738,965 23 November 2005 (23.11.2005) US

(71) Applicants (*for all designated States except US*): **APPLERA CORPORATION** [US/US]; c/o CELERA GENOMICS, Legal Patent Department, 45 West Gude Drive, C1-316, Rockville, Maryland 20850 (US). **SEATTLE GENETICS, INC.** [US/US]; 21823 30th Drive SE, Bothell, Washington 98021 (US).

(72) Inventors; and

(75) Inventors/Applicants (*for US only*): **VAN ORDEN, Karen** [US/US]; c/o CELERA GENOMICS, 45 West Gude Drive, Legal Patent Department, C1-316, Rockville, Maryland 20850 (US). **SMITH, Maria, Leia** [US/US]; c/o SEATTLE GENETICS, 21823 30th Drive SE, Bothell, Washington 9801 (US). **NESTEROVA, Albina** [US/US]; c/o SEATTLE GENETICS, INC., 21823 30th Drive SE, Bothell, Washington 98021 (US). **LEE, Candy** [GB/US]; c/o CELERA GENOMICS, 45 West Gude Drive, Legal Patent Department, C1-316, Rockville, Maryland 20850 (US). **DOMON, Bruno** [CH/CH]; c/o CELERA GENOMICS, 45 West Gude Drive, Legal Patent Department, C1-316, Rockville, Maryland 20850 (US). **RUBEN, Steve** [US/US]; c/o CELERA GENOMICS, 45 West Gude Drive, Legal Patent Department, C1-316, Rockville, Maryland 20850 (US).

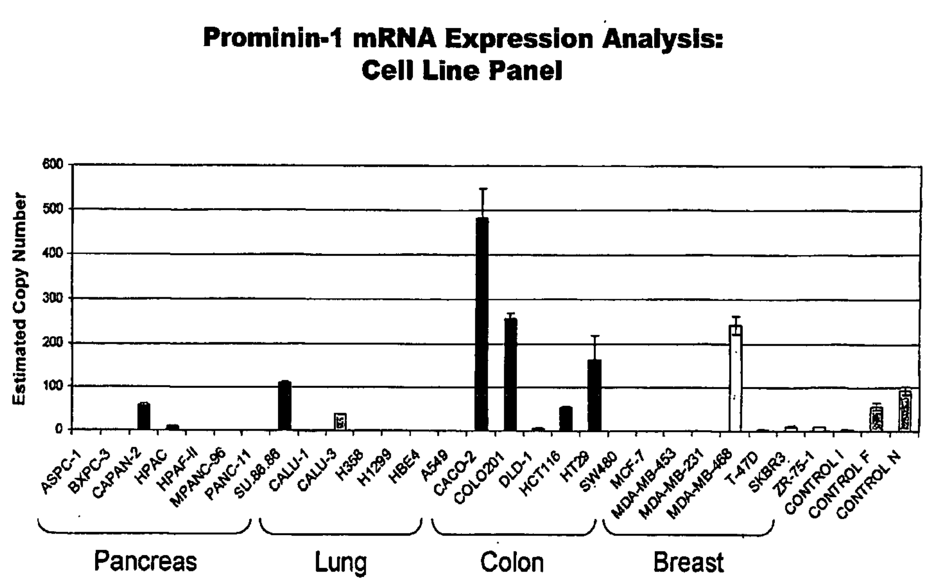
(74) Common Representative: **APPLERA CORPORATION**; c/o CELERA GENOMICS, Victor K. Lee, 45 West Gude Drive, C1-316, Rockville, Maryland 20850 (US).

(81) Designated States (*unless otherwise indicated, for every kind of national protection available*): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LV, LY, MA, MD, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, SV, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States (*unless otherwise indicated, for every kind of regional protection available*): ARIPO (BW, GH,

[Continued on next page]

(54) Title: METHODS AND COMPOSITIONS FOR TREATING DISEASES TARGETING HUMAN PROMININ-1(CD133)



(57) Abstract: Methods and compositions for detecting and treating diseases, especially cancer, and particularly breast, bladder, colon, gastrointestinal, kidney, liver, lung, melanoma, ovary, pancreatic, pharyngeal, prostate cancer and renal, associated with differential expression of prominin-1 (CD133) in disease cells compared to healthy cells. Also provided are antagonists or agonists of prominin-1, and methods for screening agents that modulate the prominin-1 level or activity in vivo or in vitro.

WO 2007/062138 A2

GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IS, IT, LT, LU, LV, MC, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

METHODS AND COMPOSITIONS FOR TREATING DISEASES TARGETING PROMININ-1 (CD133)

CONTINUITY

[0001] This application claims the benefit of U.S. Provisional Application No. 60/738,965, filed November 23, 2005; the disclosure of which is incorporated herein by reference in its entirety.

FIELD OF THE INVENTION

[0002] This invention relates to the fields of molecular biology and oncology. Specifically, the invention provides molecular markers and therapeutic agents for use in the diagnosis and treatment of diseases, especially cancer, such as for example breast, bladder, colon, gastrointestinal, glioblastoma, kidney, liver, lung, melanoma, ovary, pancreatic, pharyngeal, prostate cancer and renal.

BACKGROUND OF THE INVENTION

[0003] Cancer currently constitutes the second most common cause of death in the United States, and cancer is difficult to diagnose and treat effectively. Accordingly, there is a need in the art for improved methods for detecting and treating various cancers. The present invention fulfills these needs and further provides other related advantages, such as uses related to the treatment of other diseases.

[0004] One promising method for early diagnosis of various forms of cancer is the identification of specific biochemical moieties, termed targets, expressed differentially in cancerous cells. The targets may be either cell surface proteins, cytosolic proteins, or secreted proteins. Antibodies or other biomolecules or small molecules that will specifically recognize and bind to the targets in the cancerous cells potentially provide powerful tools for the diagnosis and treatment of the particular malignancy.

[0005] Prominin-1 (interchangeably referred to as CD133) is a member of a novel family of 5-transmembrane glycoproteins with an extracellular domain that is typically greater than 200-amino acids. An isoform of CD133, as a result of alternative splicing, contains a nine amino acid deletion in the N-terminal extracellular domain (Yu et al., 2002, J. Biol.

Chem. 23: 20711). Prominin-1 is typically expressed at subdomains of the cell surface. In hematopoietic progenitor cells, prominin-1 is enriched in plasma membrane protrusions. In epithelial cells, prominin-1 is located on the microvilli. Prominin-1 specifically interacts with cholesterol, and it has been shown that it is this interaction that allows Prominin-1 to be retained in the microvilli. Defects in prominin-1 have been identified as the cause of an autosomal recessive form of retinal degeneration. Prominin-1 has been found on tumor cells of acute myeloid leukemia (Miraglia, S. et al., 1997, *Blood* 90: 5013; Lee, S. et al., 2001, *Leuk. Res.* 25: 757), kidney cancer (Florek, M. et al., 2005, *Cell Tissue Res.* 319: 15), and non-small cell lung cancer (Hilbe, W. et al., 2004, *J. Clin. Pathol.* 57: 965). Prominin-1 is also expressed on normal bone marrow cells (Miraglia, S. et al., 1997, *Blood* 90: 5013).

[0006] Various types of stem and progenitor cells express prominin-1, including hematopoietic cells (Yin et al., *Blood* 90:5002-5012 (1997); Summers Y et al., *Stem Cells* 22:704-715 (2004)), brain cancer stem cells (Singh S. et al., *Nature* 432:396-401 (2004); Singh S. et al., *Cancer Res.* 63:5821-5828 (2003)), prostatic epithelial stem cells (Richardson G. et al., *J. Cell Sci.* 117:3539-3545 (2004)), renal progenitor cells (Bussolati B. et al., *Am. J. Path.* 166:545-555 (2005)) and endothelial progenitor cells (Peichev M. et al., *Blood* 95:952-058 (2000)).

[0007] There remains, however, a need for methods of treating prominin-1 expressing diseases, such as cancer.

SUMMARY OF THE INVENTION

[0008] A diseased, e.g., malignant, cell often differs from a normal cell by a differential expression or antigenicity of one or more proteins. These proteins, and suitable fragments thereof, are useful as markers for the diagnosis and treatment of the disease.

[0009] Based on the finding that prominin-1 is differentially expressed in disease cells, especially cancer, in comparison to normal cells, the present invention provides methods and compositions for treating diseases, especially cancer, and particularly breast, bladder, colon, gastrointestinal (e.g., gastric), glioblastoma, kidney, liver, lung, melanoma, ovary, pancreatic, pharyngeal, prostate cancer and renal, using prominin-1 as a target.

[0010] In the context of the present invention, differentially expressed prominin-1 proteins (exemplary prominin-1 protein sequences are shown in SEQ ID NOS:1-5) and suitable fragments thereof, and nucleic acids encoding said proteins (exemplary prominin-1 nucleic acid transcript sequences are shown in SEQ ID NO:6-10) and suitable fragments thereof, are referred to herein as prominin-1 protein, prominin-1 peptides, or prominin-1 nucleic acids, and collectively as prominin-1. Prominin-1 is interchangeably referred to as CD133.

[0011] The prominin-1 proteins of the present invention may serve as targets for one or more classes of therapeutic agents, including antibody therapeutics. Prominin-1 proteins of the present invention are useful in providing a target for diagnosing a disease, or predisposition to a disease expressing the protein, particularly cancer. Accordingly, the invention provides methods for detecting the presence, or levels of, a prominin-1 protein of the present invention in a biological sample such as tissues, cells and biological fluids isolated from a subject.

[0012] The diagnosis methods may detect prominin-1 nucleic acids, protein, peptides and fragments thereof that are differentially expressed in diseases in a test sample, preferably in a biological sample.

[0013] Further embodiments include, but is not limited to, monitoring the disease prognosis (recurrence), diagnosing disease stage, preventing the disease, and treating the disease.

[0014] Accordingly, the present invention provides a method for diagnosing or detecting a disease (particularly cancer) in a subject comprising: determining the level of prominin-1 in a test sample from said subject, wherein a differential level of said prominin-1 in said sample relative to the level in a control sample from a healthy subject, or the level established for a healthy subject, is indicative of the disease. The test sample includes, but is not limited to, a biological sample such as tissue, blood, serum or biological fluid.

[0015] The diagnostic method of the present invention may be suitable for monitoring disease progression or treatment progress, for example.

[0016] The diagnostic methods of the present invention may be suitable for other types of cancer, particularly other epithelial-cell related cancers.

[0017] The present invention further provides antagonists to prominin-1 protein or peptides and pharmaceutical compositions that comprise the antagonist and a suitable carrier. The antagonists may be used for treating the disease. Preferably, the antagonist is an antibody that specifically binds to a prominin-1 protein or peptide. The antibody may be used alone or in combination with another therapeutic agent (e.g., as an antibody drug conjugate or a combination therapy). In another preferred embodiment, the antagonist may be a small molecule that inhibits the function or levels of prominin-1, or an inhibitory nucleic acid molecule, such as an RNAi or antisense molecule against a prominin-1 nucleic acid.

[0018] The present invention provides additionally a pharmaceutical composition comprising an antagonist to prominin-1 of the present invention, and a pharmaceutically acceptable excipient, for treating a disease, particularly cancer.

[0019] The present invention further provides a method for treating a disease, particularly cancer, comprising administering to a patient in need of said treatment a therapeutically effective amount of the pharmaceutical composition. Such a pharmaceutical composition can include, for example, a small molecule that inhibits the function or levels of prominin-1, an inhibitory nucleic acid molecule, such as an RNAi or antisense molecule against a prominin-1 nucleic acid, an antibody or an antibody drug conjugate.

[0020] The present invention further provides a method for screening for agents that modulate prominin-1 protein activity, comprising the steps of (i) contacting a candidate agent with a prominin-1 protein, and (ii) assaying for prominin-1 protein activity, wherein a change in said activity in the presence of said agent relative to prominin-1 protein activity in the absence of said agent indicates said agent modulates said prominin-1 protein activity. Candidate agents include, but are not limited to, proteins, peptides, antibodies, nucleic acids (such as antisense RNA and RNAi fragments), and small molecules. RNAi is particularly effective at suppressing gene expression, and is therefore useful for blocking or limiting production of the prominin-1 protein, such as for treating cancer or other diseases.

[0021] The screening method may also determine a candidate agent's ability to modulate the expression level of a prominin-1 protein or nucleic acid. The method can comprise, for example, (i) contacting a candidate agent with a system that is capable of expressing a prominin-1 protein or prominin-1 mRNA, (ii) assaying for the level of a prominin-1 protein or a prominin-1 mRNA, wherein a specific change in said level in the presence of said agent relative to the level in the absence of said agent indicates said agent modulates said prominin-1 level.

[0022] The present invention further provides a method to screen for agents that bind to the prominin-1 protein, comprising the steps of (i) contacting a test agent with a prominin-1 protein and (ii) measuring the level of binding of the agent to said prominin-1 protein.

[0023] The invention will best be understood by reference to the following detailed description of the exemplary embodiments, taken in conjunction with the accompanying drawings, figures, and schemes. The discussion below is descriptive, illustrative and exemplary and is not to be taken as limiting the scope defined by any appended claims. The recitation of any reference in this application is not an admission that the reference is prior art to this application.

DESCRIPTION OF THE SEQUENCE LISTING

[0024] The Sequence Listing discloses exemplary prominin-1 protein and nucleic acid sequences. Specifically, SEQ ID NOS:1-5 of the Sequence Listing disclose the amino acid sequences of exemplary prominin-1 proteins, and SEQ ID NOS:6-10 of the Sequence Listing disclose the nucleic acid sequences of exemplary prominin-1 transcripts that encode these prominin-1 proteins. The Sequence Listing is hereby incorporated by reference pursuant to 37 CFR 1.77(b)(11).

DESCRIPTION OF THE FIGURES

[0025] **Figure 1.** Expression of prominin-1 mRNA in a panel of pancreas, lung, colon, and breast cell lines.

[0026] **Figure 2.** Expression of prominin-1 mRNA in a panel of kidney, stomach, prostate, liver, and skin cell lines.

[0027] **Figure 3.** Overexpression of prominin-1 in multiple tumor types, as indicated by immunohistochemistry (IHC).

[0028] **Figure 4.** siRNA knockdown of CD133 inhibits proliferation of HT29 colon cancer cells. **(A)** Q-PCR measuring CD133 mRNA levels demonstrates knockdown of message with three independent siRNA duplexes at 100 nM 24 h post-transfection. **(B)** Flow cytometry analysis of siRNA transfected cells indicates knockdown of CD133 protein using 100 nM CD133 siRNA (black line) compared to scrambled negative control siRNA (gray line). **(C)** Titration of CD133 siRNA (Δ) from 100 nM to 1 nM causes a dose dependent decrease in cell growth. Cell proliferation was measured using Alamar blue. Data is plotted as percent of the scrambled negative control (X). The positive control, CHK1 siRNA, is also plotted on the graph (•).

[0029] **Figure 5.** Exemplary mRNA sequence of prominin-1 (CD133) (SEQ ID NO:11), indicating siRNA target regions.

[0030] **Figure 6.** Analysis of CD133 over-expression on colon tumor cells by flow cytometry. **(A)** Normal colon and colorectal tumor samples were analyzed for CD133 expression using PE-conjugated Ac133. Results are graphed as the percent viable, EpCam+ cells expressing CD133. Tumors are organized according to stage. Representative dot plots are provided below the graph and correspond to circled data points on the graph. **(B)** CD133 epitope copy number was determined by quantitative flow cytometry. Hematopoietic cell types from bone marrow and PBMC were compared to colon tumor and matched normal adjacent tissue.

[0031] **Figure 7.** (A) Quantitative flow cytometric analysis of CD133 expression in pancreatic, hepatocellular, and gastric carcinoma cell lines. **(B-E)** Antibody-drug conjugates(ADC) targeting CD133 have potent cytotoxic activity against antigen-positive hepatocellular carcinoma cells, Hep3B. Activity against Su86.86 pancreatic cells is limited in this example. **(B)** Cytotoxicity measured by rezasurin dye conversion in Hep3B. Cells were grown in 96 well plates and treated with anti-CD133 drug conjugates for 96 hours. **(C)** Cytotoxicity measured by rezasurin dye conversion in Su86.86. Cells were treated with anti-CD133 drug conjugates for 120 hours. **(D)** Hep3B cell proliferation measured by ^{3}H-thymidine uptake after 96 hours incubation with anti-CD133 drug conjugates. **(E)** Su86.86 cell proliferation measured by ^{3}H-thymidine uptake after 120 hours incubation with anti-CD133 drug conjugates.

[0032] **Figure 8.** Effect on cell viability and cell proliferation of CD133-expressing cells treated with an anti-CD133 antibody conjugated with an auristatin derivative. AC133 antibody was conjugated to MMAE or MMAF with or without a cleavable valine-citruline (vc) linker. Viability was measured using a dye-conversion assay. Cell proliferation was measured using a ^{3}H-thymidine incorporation assay. (A) Cell viability of untransfected HEK293 cells and HEK293 cells stably expressing CD133. **(B)** Cell viability and **(C)** cell proliferation of Caco-2 cells. **(D)** Proliferation of HCT116 cells. In **(B-D)** OKT9vcMMAF (anti-transferrin receptor ADC) was included as a positive control.

[0033] **Figure 9.** (A) FACS analysis of CD133 expression in Hep3B tumor xenografts grown in SCID mice **(B)** In vivo efficacy of AC133-vcMMAF in Hep3B subcutaneous tumors in SCID mice.

[0034] **Figure 10.** Cell Transformation Assay: Soft Agar Colony Formation.

DEFINITIONS

[0035] Unless stated otherwise, the following terms and phrases as used herein are intended to have the following meanings. When trade names are used herein, applicants intend to independently include the trade name product formulation, the generic drug, and the active pharmaceutical ingredient(s) of the trade name product.

[0036] The term "alkyl" refers to a C_1-C_{18} hydrocarbon containing normal, secondary, tertiary or cyclic carbon atoms. Examples are methyl (Me, $-CH_3$), ethyl (Et, $-CH_2CH_3$), 1-propyl (n-Pr, n-propyl, $-CH_2CH_2CH_3$), 2-propyl (i-Pr, i-propyl, $-CH(CH_3)_2$), 1-butyl (n-Bu, n-butyl, $-CH_2CH_2CH_2CH_3$), 2-methyl-1-propyl (i-Bu, i-butyl, $-CH_2CH(CH_3)_2$), 2-butyl (s-Bu, s-butyl, $-CH(CH_3)CH_2CH_3$), 2-methyl-2-propyl (t-Bu, t-butyl, $-C(CH_3)_3$), 1-pentyl (n-pentyl, $-CH_2CH_2CH_2CH_2CH_3$), 2-pentyl ($-CH(CH_3)CH_2CH_2CH_3$), 3-pentyl ($-CH(CH_2CH_3)_2$), 2-methyl-2-butyl ($-C(CH_3)_2CH_2CH_3$), 3-methyl-2-butyl ($-CH(CH_3)CH(CH_3)_2$), 3-methyl-1-butyl ($-CH_2CH_2CH(CH_3)_2$), 2-methyl-1-butyl ($-CH_2CH(CH_3)CH_2CH_3$), 1-hexyl ($-CH_2CH_2CH_2CH_2CH_2CH_3$), 2-hexyl ($-CH(CH_3)CH_2CH_2CH_2CH_3$), 3-hexyl ($-CH(CH_2CH_3)(CH_2CH_2CH_3)$), 2-methyl-2-pentyl ($-C(CH_3)_2CH_2CH_2CH_3$), 3-methyl-2-pentyl ($-CH(CH_3)CH(CH_3)CH_2CH_3$), 4-methyl-2-pentyl ($-CH(CH_3)CH_2CH(CH_3)_2$), 3-methyl-3-pentyl ($-C(CH_3)(CH_2CH_3)_2$), 2-methyl-3-pentyl ($-CH(CH_2CH_3)CH(CH_3)_2$), 2,3-dimethyl-2-butyl ($-C(CH_3)_2CH(CH_3)_2$), and 3,3-dimethyl-2-butyl ($-CH(CH_3)C(CH_3)_3$.

[0037] The term "alkenyl" refers to a C_2-C_{18} hydrocarbon containing normal, secondary, tertiary or cyclic carbon atoms with at least one site of unsaturation, i.e., a carbon-carbon, sp^2 double bond. Examples include, but are not limited to: ethylene or vinyl ($-CH=CH_2$), allyl ($-CH_2CH=CH_2$), cyclopentenyl ($-C_5H_7$), and 5-hexenyl ($-CH_2\ CH_2CH_2CH_2CH=CH_2$).

[0038] The term "alkynyl" refers to a C_2-C_{18} hydrocarbon containing normal, secondary, tertiary or cyclic carbon atoms with at least one site of unsaturation, i.e., a carbon-carbon, sp triple bond. Examples include, but are not limited to: acetylenic ($-C\equiv CH$) and propargyl ($-CH_2C\equiv CH$).

[0039] The term "alkylene" refers to a saturated, branched or straight chain or cyclic hydrocarbon radical of 1-18 carbon atoms, and having two monovalent radical centers derived by the removal of two hydrogen atoms from the same or two different carbon atoms of a parent alkane. Typical alkylenes include, but are not limited to: methylene ($-CH_2-$) 1,2-ethyl ($-CH_2CH_2-$), 1,3-propyl ($-CH_2CH_2CH_2-$), 1,4-butyl ($-CH_2CH_2CH_2CH_2-$), and the like.

[0040] The term "alkenylene" refers to an unsaturated, branched or straight chain or cyclic hydrocarbon radical of 2-18 carbon atoms, and having two monovalent radical centers derived by the removal of two hydrogen atoms from the same or two different carbon atoms of a parent alkene. Typical alkenylene radicals include, but are not limited to: 1,2-ethylene (-CH=CH-).

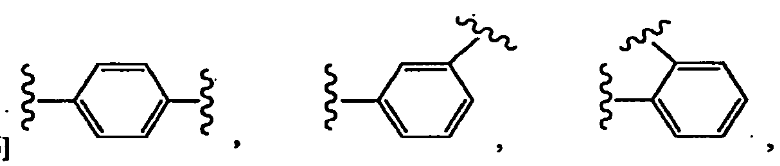
[0041] The term "alkynylene" refers to an unsaturated, branched or straight chain or cyclic hydrocarbon radical of 2-18 carbon atoms, and having two monovalent radical centers derived by the removal of two hydrogen atoms from the same or two different carbon atoms of a parent alkyne. Typical alkynylene radicals include, but are not limited to: acetylene (-C≡C-), propargyl ($-CH_2C{\equiv}C-$), and 4-pentynyl ($-CH_2CH_2CH_2C{\equiv}CH-$).

[0042] The term "aryl" refers to a monovalent aromatic hydrocarbon radical of 6-20 carbon atoms derived by the removal of one hydrogen atom from a single carbon atom of a parent aromatic ring system. Some aryl groups are represented in the exemplary structures as "Ar". An aryl group can be unsubstituted or substituted. Typical aryl groups include, but are not limited to, radicals derived from benzene, substituted benzene, phenyl, naphthalene, anthracene, biphenyl, and the like. A carbocyclic aromatic group or a heterocyclic aromatic group can be unsubstituted or substituted with one or more groups including, but not limited to, $-C_1-C_8$ alkyl, $-O-(C_1-C_8$ alkyl), -aryl, -C(O)R', -OC(O)R', -C(O)OR', -C(O)NH_2, -C(O)NHR', -C(O)N$(R')_2$ -NHC(O)R', $-S(O)_2R'$, -S(O)R', -OH, -halogen, $-N_3$, $-NH_2$, -NH(R'), $-N(R')_2$ and -CN; wherein each R' is independently selected from H, $-C_1-C_8$ alkyl and aryl.

[0043] The term "arylalkyl" refers to an acyclic alkyl radical in which one of the hydrogen atoms bonded to a carbon atom, typically a terminal or sp^3 carbon atom, is replaced with an aryl radical. Typical arylalkyl groups include, but are not limited to, benzyl, 2-phenylethan-1-yl, 2-phenylethen-1-yl, naphthylmethyl, 2-naphthylethan-1-yl, 2-naphthylethen-1-yl, naphthobenzyl, 2-naphthophenylethan-1-yl and the like. The arylalkyl group comprises 6 to 20 carbon atoms, e.g., the alkyl moiety, including alkanyl, alkenyl or alkynyl groups, of the arylalkyl group is 1 to 6 carbon atoms and the aryl moiety is 5 to 14 carbon atoms.

[0044] The term "heteroarylalkyl" refers to an acyclic alkyl radical in which one of the hydrogen atoms bonded to a carbon atom, typically a terminal or sp^3 carbon atom, is replaced with a heteroaryl radical. Typical heteroarylalkyl groups include, but are not limited to, 2-benzimidazolylmethyl, 2-furylethyl, and the like. The heteroarylalkyl group comprises 6 to 20 carbon atoms, e.g., the alkyl moiety, including alkanyl, alkenyl or alkynyl groups, of the heteroarylalkyl group is 1 to 6 carbon atoms and the heteroaryl moiety is 5 to 14 carbon atoms and 1 to 3 heteroatoms selected from N, O, P, and S. The heteroaryl moiety of the heteroarylalkyl group may be a monocycle having 3 to 7 ring members (2 to 6 carbon atoms and 1 to 3 heteroatoms selected from N, O, P and S) or a bicycle having 7 to 10 ring members (4 to 9 carbon atoms and 1 to 3 heteroatoms selected from N, O, P, and S), for example: a bicyclo [4,5], [5,5], [5,6], or [6,6] system.

[0045] The term "arylene" refers to an aryl group which has two covalent bonds and can be in the ortho, meta, or para configurations as shown in the following structures:

[0046] 

[0047] in which the phenyl group can be unsubstituted or substituted with up to four groups including, but not limited to, -C_1-C_8 alkyl, -O-(C_1-C_8 alkyl), -aryl, -C(O)R', -OC(O)R',
-C(O)OR', -C(O)NH_2 , -C(O)NHR', -C(O)N(R')$_2$ -NHC(O)R', -S(O)$_2$R', -S(O)R', -OH, -halogen, -N_3 , -NH_2, -NH(R'), -N(R')$_2$ and -CN; wherein each R' is independently selected from H, -C_1-C_8 alkyl and aryl.

[0048] The terms "substituted alkyl", "substituted aryl", and "substituted arylalkyl" refer to alkyl, aryl, and arylalkyl, respectively, in which one or more hydrogen atoms are each independently replaced with a substituent. Typical substituents include, but are not limited to, -X, -R, -O^-, -OR, -SR, -S^-, -NR_2, -NR_3, =NR, -CX_3, -CN, -OCN, -SCN, -N=C=O, -NCS, -NO, -NO_2, =N_2, -N_3, NC(=O)R, -C(=O)R, -C(=O)NR_2, -SO_3^-, -SO_3H, -S(=O)$_2$R, -OS(=O)$_2$OR, -S(=O)$_2$NR, -S(=O)R, -OP(=O)(OR)$_2$, -P(=O)(OR)$_2$, -PO^-_3, -PO_3H_2, -C(=O)R, -C(=O)X, -C(=S)R, -CO_2R, -CO_2^-, -C(=S)OR, -C(=O)SR, -C(=S)SR, -C(=O)NR_2,

-C(=S)NR_2, or -C(=NR)NR_2, where each X is independently a halogen: F, Cl, Br, or I; and each R is independently -H, C_2-C_{18} alkyl, C_6-C_{20} aryl, C_3-C_{14} heterocycle, a protecting group or a prodrug moiety. Alkylene, alkenylene, and alkynylene groups as described above may also be similarly substituted.

[0049] The terms "heteroaryl" and "heterocycle" refer to a ring system in which one or more ring atoms is a heteroatom, e.g., nitrogen, oxygen, and sulfur. The heterocycle radical comprises 1 to 20 carbon atoms and 1 to 3 heteroatoms selected from N, O, P, and S. A heterocycle may be a monocycle having 3 to 7 ring members (2 to 6 carbon atoms and 1 to 3 heteroatoms selected from N, O, P, and S) or a bicycle having 7 to 10 ring members (4 to 9 carbon atoms and 1 to 3 heteroatoms selected from N, O, P, and S), for example: a bicyclo [4,5], [5,5], [5,6], or [6,6] system. Heterocycles are described in Paquette, "Principles of Modern Heterocyclic Chemistry" (W.A. Benjamin, New York, 1968), particularly Chapters 1, 3, 4, 6, 7, and 9; "The Chemistry of Heterocyclic Compounds, A series of Monographs" (John Wiley & Sons, New York, 1950 to present), in particular Volumes 13, 14, 16, 19, and 28; and J. Am. Chem. Soc. (1960) 82:5566.

[0050] Examples of heterocycles include by way of example and not limitation pyridyl, dihydroypyridyl, tetrahydropyridyl (piperidyl), thiazolyl, tetrahydrothiophenyl, sulfur oxidized tetrahydrothiophenyl, pyrimidinyl, furanyl, thienyl, pyrrolyl, pyrazolyl, imidazolyl, tetrazolyl, benzofuranyl, thianaphthalenyl, indolyl, indolenyl, quinolinyl, isoquinolinyl, benzimidazolyl, piperidinyl, 4-piperidonyl, pyrrolidinyl, 2-pyrrolidonyl, pyrrolinyl, tetrahydrofuranyl, bis-tetrahydrofuranyl, tetrahydropyranyl, bis-tetrahydropyranyl, tetrahydroquinolinyl, tetrahydroisoquinolinyl, decahydroquinolinyl, octahydroisoquinolinyl, azocinyl, triazinyl, 6H-1,2,5-thiadiazinyl, 2H,6H-1,5,2-dithiazinyl, thienyl, thianthrenyl, pyranyl, isobenzofuranyl, chromenyl, xanthenyl, phenoxathinyl, 2H-pyrrolyl, isothiazolyl, isoxazolyl, pyrazinyl, pyridazinyl, indolizinyl, isoindolyl, 3H-indolyl, 1H-indazolyl, purinyl, 4H-quinolizinyl, phthalazinyl, naphthyridinyl, quinoxalinyl, quinazolinyl, cinnolinyl, pteridinyl, 4aH-carbazolyl, carbazolyl, β-carbolinyl, phenanthridinyl, acridinyl, pyrimidinyl, phenanthrolinyl, phenazinyl, phenothiazinyl, furazanyl, phenoxazinyl, isochromanyl, chromanyl, imidazolidinyl, imidazolinyl,

pyrazolidinyl, pyrazolinyl, piperazinyl, indolinyl, isoindolinyl, quinuclidinyl, morpholinyl, oxazolidinyl, benzotriazolyl, benzisoxazolyl, oxindolyl, benzoxazolinyl, and isatinoyl.

[0051] By way of example and not limitation, carbon-bonded heterocycles are bonded at the following positions: position 2, 3, 4, 5, or 6 of a pyridine; position 3, 4, 5, or 6 of a pyridazine; position 2, 4, 5, or 6 of a pyrimidine; position 2, 3, 5, or 6 of a pyrazine; position 2, 3, 4, or 5 of a furan, tetrahydrofuran, thiofuran, thiophene, pyrrole or tetrahydropyrrole; position 2, 4, or 5 of an oxazole, imidazole or thiazole; position 3, 4, or 5 of an isoxazole, pyrazole, or isothiazole; position 2 or 3 of an aziridine; position 2, 3, or 4 of an azetidine; position 2, 3, 4, 5, 6, 7, or 8 of a quinoline; or position 1, 3, 4, 5, 6, 7, or 8 of an isoquinoline. Still more typically, carbon bonded heterocycles include 2-pyridyl, 3-pyridyl, 4-pyridyl, 5-pyridyl, 6-pyridyl, 3-pyridazinyl, 4-pyridazinyl, 5-pyridazinyl, 6-pyridazinyl, 2-pyrimidinyl, 4-pyrimidinyl, 5-pyrimidinyl, 6-pyrimidinyl, 2-pyrazinyl, 3-pyrazinyl, 5-pyrazinyl, 6-pyrazinyl, 2-thiazolyl, 4-thiazolyl, or 5-thiazolyl.

[0052] By way of example and not limitation, nitrogen bonded heterocycles are bonded at position 1 of an aziridine, azetidine, pyrrole, pyrrolidine, 2-pyrroline, 3-pyrroline, imidazole, imidazolidine, 2-imidazoline, 3-imidazoline, pyrazole, pyrazoline, 2-pyrazoline, 3-pyrazoline, piperidine, piperazine, indole, indoline, or 1H-indazole; position 2 of a isoindole, or isoindoline; position 4 of a morpholine; and position 9 of a carbazole, or β-carboline. Still more typically, nitrogen bonded heterocycles include 1-aziridyl, 1-azetedyl, 1-pyrrolyl, 1-imidazolyl, 1-pyrazolyl, and 1-piperidinyl.

[0053] The term "carbocycle" refers to a saturated or unsaturated ring having 3 to 7 carbon atoms as a monocycle or 7 to 12 carbon atoms as a bicycle. Monocyclic carbocycles have 3 to 6 ring atoms, still more typically 5 or 6 ring atoms. Bicyclic carbocycles have 7 to 12 ring atoms, e.g., arranged as a bicyclo [4,5], [5,5], [5,6] or [6,6] system, or 9 or 10 ring atoms arranged as a bicyclo [5,6] or [6,6] system. Examples of monocyclic carbocycles include cyclopropyl, cyclobutyl, cyclopentyl, 1-cyclopent-1-enyl, 1-cyclopent-2-enyl, 1-cyclopent-3-enyl, cyclohexyl, 1-cyclohex-1-enyl, 1-cyclohex-2-enyl, 1-cyclohex-3-enyl, cycloheptyl, and cyclooctyl.

[0054] The term "C_1-C_8 alkyl," as used herein refers to a straight chain or branched, saturated or unsaturated hydrocarbon having from 1 to 8 carbon atoms. Representative "C_1-C_8 alkyl" groups include, but are not limited to, -methyl, -ethyl, -n-propyl, -n-butyl, -n-pentyl, -n-hexyl, -n-heptyl, and -n-octyl,; while branched C_1-C_8 alkyls include, but are not limited to, -isopropyl, -sec-butyl, -isobutyl, -tert-butyl, -isopentyl, 2-methylbutyl, isohexyl, 2-methylpentyl, 3-methylpentyl, 2,2-dimethylbutyl, 2,3-dimethylbutyl, 2,2-dimethylpentyl, 2,3-dimethylpentyl, 3,3-dimethylpentyl, 2,3,4-trimethylpentyl, 3-methylhexyl, 2,2-dimethylhexyl, 2,4-dimethylhexyl, 2,5-dimethylhexyl, 3,5-dimethylhexyl, 2,4-dimethylpentyl, isoheptyl, isooctyl, 2-methylheptyl, 3-methylheptyl; unsaturated C_1-C_8 alkyls include, but are not limited to, -vinyl, -allyl, -1-butenyl, -2-butenyl, -isobutylenyl, -1-pentenyl, -2-pentenyl, -3-methyl-1-butenyl, -2-methyl-2-butenyl, -2,3-dimethyl-2-butenyl, 1-hexyl, 2-hexyl, 3-hexyl,-acetylenyl, -propynyl, -1-butynyl, -2-butynyl, -1-pentynyl, -2-pentynyl, and -3-methyl-1 butynyl. A C_1-C_8 alkyl group can be unsubstituted or substituted with one or more groups including, but not limited to, -C_1-C_8 alkyl, -O-(C_1-C_8 alkyl), -aryl, -C(O)R', -OC(O)R', -C(O)OR', -C(O)NH_2 , -C(O)NHR', -C(O)N$(R')_2$ -NHC(O)R', -SO_3R', -S$(O)_2$R', -S(O)R', -OH, -halogen, -N_3 , -NH_2, -NH(R'), -N$(R')_2$ and -CN; where each R' is independently selected from H, -C_1-C_8 alkyl and aryl.

[0055] A "C_3-C_8 carbocycle" is a 3-, 4-, 5-, 6-, 7- or 8-membered saturated or unsaturated non-aromatic carbocyclic ring. Representative C_3-C_8 carbocycles include, but are not limited to, -cyclopropyl, -cyclobutyl, -cyclopentyl, -cyclopentadienyl, -cyclohexyl, -cyclohexenyl, -1,3-cyclohexadienyl, -1,4-cyclohexadienyl, -cycloheptyl, -1,3-cycloheptadienyl, -1,3,5-cycloheptatrienyl, -cyclooctyl, and -cyclooctadienyl. A C_3-C_8 carbocycle group can be unsubstituted or substituted with one or more groups including, but not limited to, -C_1-C_8 alkyl, -O-(C_1-C_8 alkyl), -aryl, -C(O)R', -OC(O)R', -C(O)OR', -C(O)NH_2 , -C(O)NHR', -C(O)N$(R')_2$, -NHC(O)R', -S$(O)_2$R', -S(O)R', -OH, -halogen, -N_3 , -NH_2, -NH(R'), -N$(R')_2$ and -CN; where each R' is independently selected from H, -C_1-C_8 alkyl and aryl.

[0056] A "C_3-C_8 carbocyclo" refers to a C_3-C_8 carbocycle group defined above wherein one of the carbocycle groups' hydrogen atoms is replaced with a bond.

[0057] A "C_1-C_{10} alkylene" is a straight chain, saturated hydrocarbon group of the formula -$(CH_2)_{1\text{-}10}$-. Examples of a C_1-C_{10} alkylene include methylene, ethylene, propylene, butylene, pentylene, hexylene, heptylene, ocytylene, nonylene and decalene.

[0058] A "C_3-C_8 heterocycle" refers to an aromatic or non-aromatic C_3-C_8 carbocycle in which one to four of the ring carbon atoms are independently replaced with a heteroatom from the group consisting of O, S and N. Representative examples of a C_3-C_8 heterocycle include, but are not limited to, benzofuranyl, benzothiophene, indolyl, benzopyrazolyl, coumarinyl, isoquinolinyl, pyrrolyl, thiophenyl, furanyl, thiazolyl, imidazolyl, pyrazolyl, triazolyl, quinolinyl, pyrimidinyl, pyridinyl, pyridonyl, pyrazinyl, pyridazinyl, isothiazolyl, isoxazolyl and tetrazolyl. A C_3-C_8 heterocycle can be unsubstituted or substituted with up to seven groups including, but not limited to, -C_1-C_8 alkyl, -O-(C_1-C_8 alkyl), -aryl, -C(O)R', -OC(O)R', -C(O)OR', -C(O)NH_2 , -C(O)NHR', -C(O)N(R')$_2$, -NHC(O)R', -S(O)$_2$R', -S(O)R', -OH, -halogen, -N_3 , -NH_2, -NH(R'), -N(R')$_2$ and -CN; wherein each R' is independently selected from H, -C_1-C_8 alkyl and aryl.

[0059] "C_3-C_8 heterocyclo" refers to a C_3-C_8 heterocycle group defined above wherein one of the heterocycle group's hydrogen atoms is replaced with a bond. A C_3-C_8 heterocyclo can be unsubstituted or substituted with up to six groups including, but not limited to, -C_1-C_8 alkyl, -O-(C_1-C_8 alkyl), -aryl, -C(O)R', -OC(O)R', -C(O)OR', -C(O)NH_2 , -C(O)NHR', -C(O)N(R')$_2$ -NHC(O)R', -S(O)$_2$R', -S(O)R', -OH, -halogen, -N_3 , -NH_2, -NH(R'), -N(R')$_2$ and -CN; wherein each R' is independently selected from H, -C_1-C_8 alkyl and aryl.

[0060] The phrase "pharmaceutically acceptable salt" refers to a pharmaceutically acceptable organic or inorganic salt of a ligand drug conjugate or linker drug conjugate. The conjugates may contain at least one amino group, and accordingly acid addition salts can be formed with the amino group. Exemplary salts include, but are not limited, to sulfate, citrate, acetate, oxalate, chloride, bromide, iodide, nitrate, bisulfate, phosphate, acid phosphate, isonicotinate, lactate, salicylate, acid citrate, tartrate, oleate, tannate, pantothenate, bitartrate, ascorbate, succinate, maleate, gentisinate, fumarate, gluconate, glucuronate, saccharate, formate, benzoate, glutamate, methanesulfonate, ethanesulfonate,

benzenesulfonate, p-toluenesulfonate, and pamoate (i.e., 1,1' methylene bis -(2 hydroxy 3 naphthoate)) salts. A pharmaceutically acceptable salt may involve the inclusion of another molecule such as an acetate ion, a succinate ion or other counterion. The counterion may be any organic or inorganic moiety that stabilizes the charge on the parent compound. Furthermore, a pharmaceutically acceptable salt may have more than one charged atom in its structure. Instances where multiple charged atoms are part of the pharmaceutically acceptable salt can have multiple counter ions. Hence, a pharmaceutically acceptable salt can have one or more charged atoms and/or one or more counterion.

[0061] The phrases "pharmaceutically acceptable solvate" or "solvate" refer to an association of one or more solvent molecules and a ligand drug conjugate or linker drug conjugate. Examples of solvents that form pharmaceutically acceptable solvates include, but are not limited to, water, isopropanol, ethanol, methanol, DMSO, ethyl acetate, acetic acid, and ethanolamine.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS:

1. Prominin-1 Protein and Peptides

[0062] The present invention provides isolated prominin-1 proteins that consist of, consist essentially of, or comprise the amino acid sequence of SEQ ID NOS:1-5, respectively encoded by the nucleic acid molecules having the nucleotide sequences of SEQ ID NOS:6-10, as well as all obvious variants of these proteins that are within the art to make and use. Some of these variants are described in detail below.

[0063] A prominin-1 peptide or protein can be attached to heterologous sequences to form chimeric or fusion proteins. Such chimeric and fusion proteins comprise a peptide operatively linked to a heterologous protein having an amino acid sequence not substantially homologous to the peptide. "Operatively linked" indicates that the peptide and the heterologous protein are fused in-frame. The heterologous protein can be fused to the N-terminus or C-terminus of the peptide.

[0064] In some uses, the fusion protein does not affect the activity of the peptide or protein per se. For example, the fusion protein can include, but is not limited to, beta-galactosidase fusions, yeast two-hybrid GAL fusions, poly-His fusions, MYC-tagged, HI-tagged and Ig fusions. Such fusion proteins, particularly poly-His fusions, can facilitate the purification of recombinant prominin-1 proteins or peptides. In certain host cells (e.g., mammalian host cells), expression and/or secretion of a protein can be increased by using a heterologous signal sequence.

[0065] A chimeric or fusion prominin-1 protein or peptide can be produced by standard recombinant DNA techniques. For example, DNA fragments coding for the different protein sequences are ligated together in-frame in accordance with conventional techniques. In another embodiment, the fusion gene can be synthesized by conventional techniques including automated DNA synthesizers. Alternatively, PCR amplification of gene fragments can be carried out using anchor primers which give rise to complementary overhangs between two consecutive gene fragments which can subsequently be annealed and re-amplified to generate a chimeric gene sequence (see Ausubel et al., Current Protocols in Molecular Biology, 1992-2006). Moreover, many expression vectors are commercially available that already encode a fusion moiety (e.g., a GST protein). A prominin-1-encoding nucleic acid can be cloned into such an expression vector such that the fusion moiety is linked in-frame to the prominin-1 protein or peptide.

[0066] Variants of the prominin-1 protein can readily be identified/made using molecular techniques and the sequence information disclosed herein. Further, such variants can readily be distinguished from other peptides based on sequence and/or structural homology to the prominin-1 peptides of the present invention. The degree of homology/identity present will be based primarily on whether the peptide is a functional variant or non-functional variant, the amount of divergence present in the paralog family and the evolutionary distance between the orthologs.

[0067] To determine the percent identity of two amino acid sequences or two nucleic acid sequences, the sequences are aligned for optimal comparison purposes (e.g., gaps can be introduced in one or both of a first and a second amino acid or nucleic acid sequence for

optimal alignment and non-homologous sequences can be disregarded for comparison purposes). In a preferred embodiment, at least 30%, 40%, 50%, 60%, 70%, 80%, or 90% or more of the length of a reference sequence is aligned for comparison purposes. The amino acid residues or nucleotides at corresponding amino acid positions or nucleotide positions are then compared. When a position in the first sequence is occupied by the same amino acid residue or nucleotide as the corresponding position in the second sequence, then the molecules are identical at that position (as used herein amino acid or nucleic acid "identity" is equivalent to amino acid or nucleic acid "homology"). The percent identity between the two sequences is a function of the number of identical positions shared by the sequences, taking into account the number of gaps, and the length of each gap, which need to be introduced for optimal alignment of the two sequences.

[0068] The comparison of sequences and determination of percent identity and similarity between two sequences can be accomplished using a mathematical algorithm. (Computational Molecular Biology, Lesk, A.M., ed., Oxford University Press, New York, 1988; Biocomputing: Informatics and Genome Projects, Smith, D.W., ed., Academic Press, New York, 1993; Computer Analysis of Sequence Data, Part 1, Griffin, A.M., and Griffin, H.G., eds., Humana Press, New Jersey, 1994; Sequence Analysis in Molecular Biology, von Heinje, G., Academic Press, 1987; and Sequence Analysis Primer, Gribskov, M. and Devereux, J., eds., Stockton Press, New York, 1991). In a preferred embodiment, the percent identity between two amino acid sequences is determined using the Needleman and Wunsch (J. Mol. Biol. (48):444-453 (1970)) algorithm which has been incorporated into the GAP program in the GCG software package, using either a Blossom 62 matrix or a PAM250 matrix, and a gap weight of 16, 14, 12, 10, 8, 6, or 4 and a length weight of 1, 2, 3, 4, 5, or 6. In yet another preferred embodiment, the percent identity between two nucleotide sequences is determined using the GAP program in the GCG software package (Devereux, J., et al., Nucleic Acids Res. 12(1):387 (1984)), using a NWSgapdna.CMP matrix and a gap weight of 40, 50, 60, 70, or 80 and a length weight of 1, 2, 3, 4, 5, or 6. In another embodiment, the percent identity between two amino acid or nucleotide sequences is determined using the algorithm of E. Myers and W. Miller (CABIOS, 4:11-17 (1989)) which has been incorporated into the ALIGN program (version 2.0), using a PAM120 weight residue table, a gap length penalty of 12 and a gap penalty of 4.

[0069] The nucleic acids and protein sequences of the present invention can further be used as a "query sequence" to perform a search against sequence databases to, for example, identify other family members or related sequences. Such searches can be performed using the NBLAST and XBLAST programs (version 2.0) of Altschul, et al. (J. Mol. Biol. 215:403-10 (1990)). BLAST nucleotide searches can be performed with the NBLAST program, score = 100, wordlength = 12, to obtain nucleotide sequences homologous to the nucleic acid molecules of the invention. BLAST protein searches can be performed with the XBLAST program, score = 50, wordlength = 3 to obtain amino acid sequences homologous to the proteins of the invention. To obtain gapped alignments for comparison purposes, Gapped BLAST can be utilized as described in Altschul et al. (Nucleic Acids Res. 25(17):3389-3402 (1997)). When utilizing BLAST and gapped BLAST programs, the default parameters of the respective programs (e.g., XBLAST and NBLAST) can be used.

[0070] Allelic variants of a prominin-1 peptide can readily be identified as being a human protein having a high degree (significant) of sequence homology/identity to at least a portion of the prominin-1 peptide as well as being encoded by the same genetic locus as the prominin-1 peptide provided herein. The genetic locus can readily be determined based on the genomic information. As used herein, two proteins (or a region of the proteins) have significant homology (also referred to as substantial homology) when the amino acid sequences are typically at least about 70-80%, 80-90%, and more typically at least about 90-95% or more homologous. A significantly homologous amino acid sequence, according to the present invention, will be encoded by a nucleic acid sequence that will hybridize to a prominin-1 peptide encoding nucleic acid molecule under stringent conditions as more fully described below.

[0071] Paralogs of a prominin-1 peptide can readily be identified as having some degree of significant sequence homology/identity to at least a portion of the prominin-1 peptide, as being encoded by a gene from humans, and as having similar activity or function. Two proteins will typically be considered paralogs when the amino acid sequences are typically at least about 60% or greater, and more typically at least about 70% or greater homology through a given region or domain. Such paralogs will be encoded by a nucleic acid

sequence that will hybridize to a prominin-1 peptide encoding nucleic acid molecule under moderate to stringent conditions as more fully described below.

[0072] Orthologs of a prominin-1 peptide can readily be identified as having some degree of significant sequence homology/identity to at least a portion of the prominin-1 peptide as well as being encoded by a gene from another organism. Preferred orthologs will be isolated from mammals, preferably primates, for the development of human therapeutic targets and agents. Such orthologs will be encoded by a nucleic acid sequence that will hybridize to a prominin-1 peptide-encoding nucleic acid molecule under moderate to stringent conditions, as more fully described below, depending on the degree of relatedness of the two organisms yielding the proteins.

[0073] Non-naturally occurring variants of the prominin-1 peptides of the present invention can readily be generated using recombinant techniques. Such variants include, but are not limited to deletions, additions and substitutions in the amino acid sequence of the prominin-1 peptide. For example, one class of substitutions is conserved amino acid substitutions. Such substitutions are those that substitute a given amino acid in a prominin-1 peptide by another amino acid of like characteristics. Typically seen as conservative substitutions are the replacements, one for another, among the aliphatic amino acids Ala, Val, Leu, and Ile; interchange of the hydroxyl residues Ser and Thr; exchange of the acidic residues Asp and Glu; substitution between the amide residues Asn and Gln; exchange of the basic residues Lys and Arg; and replacements among the aromatic residues Phe and Tyr. Guidance concerning which amino acid changes are likely to be phenotypically silent are found in Bowie et al., Science 247:1306-1310 (1990).

[0074] Variant prominin-1 peptides can be fully functional or can lack function in one or more activities, e.g., ability to bind substrate, ability to phosphorylate substrate, ability to mediate signaling, etc. Fully functional variants typically contain only conservative variation or variation in non-critical residues or in non-critical regions.

[0075] Non-functional variants typically contain one or more non-conservative amino acid substitutions, deletions, insertions, inversions, or truncation or a substitution, insertion, inversion, or deletion in a critical residue or critical region.

[0076] Amino acids that are essential for function can be identified by methods known in the art, such as site-directed mutagenesis or alanine-scanning mutagenesis (Cunningham et al., Science 244:1081-1085 (1989)). The latter procedure introduces single alanine mutations at every residue in the molecule. The resulting mutant molecules are then tested for biological activity such as prominin-1 activity or in assays such as an in vitro proliferative activity. Sites that are critical for binding partner/substrate binding can also be determined by structural analysis such as crystallization, nuclear magnetic resonance or photoaffinity labeling (Smith et al., J. Mol. Biol. 224:899-904 (1992); de Vos et al. Science 255:306-312 (1992)).

[0077] The present invention further provides fragments of prominin-1, in addition to and peptides that comprise and consist of such fragments. As used herein, a fragment comprises at least 8, 10, 12, 14, 16, 18, 20 or more contiguous amino acid residues from prominin-1. Such fragments can be chosen based on the ability to retain one or more of the biological activities of prominin-1 or could be chosen for the ability to perform a function, e.g. bind a substrate or act as an immunogen. Particularly important fragments are biologically active fragments, peptides that are, for example, about 8 or more amino acids in length. Such fragments will typically comprise a domain or motif of prominin-1, e.g., active site, a transmembrane domain or a binding domain.. Further, possible fragments include, but are not limited to, domain or motif containing fragments, soluble peptide fragments, and fragments containing immunogenic structures. Predicted domains and functional sites are readily identifiable by computer programs well known and readily available to those of skill in the art (e.g., PROSITE analysis).

[0078] Polypeptides often contain amino acids other than the 20 amino acids commonly referred to as the 20 naturally occurring amino acids. Further, many amino acids, including the terminal amino acids, may be modified by natural processes, such as processing and other post-translational modifications, or by chemical modification techniques well known in the art. Common modifications that occur naturally in prominin-1 are described in basic texts, detailed monographs, and the research literature, and they are well known to those of skill in the art.

[0079] Known modifications include, but are not limited to, acetylation, acylation, ADP-ribosylation, amidation, covalent attachment of flavin, covalent attachment of a heme moiety, covalent attachment of a nucleotide or nucleotide derivative, covalent attachment of a lipid or lipid derivative, covalent attachment of phosphotidylinositol, cross-linking, cyclization, disulfide bond formation, demethylation, formation of covalent crosslinks, formation of cystine, formation of pyroglutamate, formylation, gamma carboxylation, glycosylation, GPI anchor formation, hydroxylation, iodination, methylation, myristoylation, oxidation, proteolytic processing, phosphorylation, prenylation, racemization, selenoylation, sulfation, transfer-RNA mediated addition of amino acids to proteins such as arginylation, and ubiquitination.

[0080] Such modifications are well known to those of skill in the art and have been described in great detail in the scientific literature. Several particularly common modifications, glycosylation, lipid attachment, sulfation, gamma-carboxylation of glutamic acid residues, hydroxylation and ADP-ribosylation, for instance, are described in most basic texts, such as Proteins - Structure and Molecular Properties, 2nd Ed., T.E. Creighton, W. H. Freeman and Company, New York (1993). Many detailed reviews are available on this subject, such as by Wold, F., Posttranslational Covalent Modification of Proteins, B.C. Johnson, Ed., Academic Press, New York 1-12 (1983); Seifter et al. (Meth. Enzymol. 182: 626-646 (1990)) and Rattan et al. (Ann. N.Y. Acad. Sci. 663:48-62 (1992)).

[0081] Accordingly, the prominin-1 protein and peptides of the present invention also encompasses derivatives or analogs in which a substituted amino acid residue is not one encoded by the genetic code, in which a substituent group is included, in which the mature prominin-1 is fused with another compound, such as a compound to increase the half-life of prominin-1 (for example, polyethylene glycol), or in which the additional amino acids are fused to the mature prominin-1, such as a leader or secretory sequence or a sequence for purification of the mature prominin-1 or a pro-protein sequence.

2. Antibodies Against Prominin-1 Protein or Fragments Thereof

[0082] Antibodies that selectively bind to the prominin-1 protein or peptides of the present invention can be made using standard procedures known to those of ordinary skills in the

art. The term "antibody" is used in the broadest sense, and specifically covers monoclonal antibodies (including full length monoclonal antibodies), polyclonal antibodies, multispecific antibodies (e.g., bispecific antibodies), chimeric antibodies, humanized antibodies, and antibody fragments (e.g., Fab, $F(ab')_2$, Fv and Fv-containing binding proteins) so long as they exhibit the desired biological activity. Antibodies (Abs) and immunoglobulins (Igs) are glycoproteins having the same structural characteristics. While antibodies exhibit binding specificity to a specific antigen, immunoglobulins include both antibodies and other antibody-like molecules that lack antigen specificity. Antibodies can be of the IgG, IgE, IgM, IgD, and IgA class or subclass thereof (e.g., IgG1, IgG2, IgG3, IgG4, IgA1 and IgA2).

[0083] As used herein, antibodies are usually heterotetrameric glycoproteins of about 150,000 daltons, composed of two identical light (L) chains and two identical heavy (H) chains (referred to as an "intact" antibody). Each light chain is linked to a heavy chain by one covalent disulfide bond, while the number of disulfide linkages varies between the heavy chains of different immunoglobulin isotypes. Each heavy and light chain also has regularly spaced intrachain disulfide bridges. Each heavy chain has at one end a variable domain (VH) followed by a number of constant domains. Each light chain has a variable domain at one end (VL) and a constant domain at its other end. The constant domain of the light chain is aligned with the first constant domain of the heavy chain, and the light chain variable domain is aligned with the variable domain of the heavy chain. Particular amino acid residues are believed to form an interface between the light and heavy chain variable domains. Chothia et al., J. Mol. Biol. 186:651-63 (1985); Novotny and Haber, Proc. Natl. Acad. Sci. USA 82:4592-4596 (1985).

[0084] An "isolated" antibody is one which has been identified and separated and/or recovered from a component of the environment in which it is produced. Contaminant components of its production environment are materials that would interfere with diagnostic or therapeutic uses for the antibody, and may include enzymes, hormones, and other proteinaceous or nonproteinaceous solutes. In preferred embodiments, the antibody will be purified as measurable by at least three different methods: 1) to greater than 95% by weight of antibody as determined by the Lowry method, and most preferably more than

99% by weight; 2) to a degree sufficient to obtain at least 15 residues of N-terminal or internal amino acid sequence by use of a spinning cup sequenator; or 3) to homogeneity by SDS-PAGE under reducing or non-reducing conditions using Coomasie blue or, preferably, silver stain. Isolated antibody includes the antibody in situ within recombinant cells since at least one component of the antibody's natural environment will not be present. Ordinarily, however, the isolated antibody will be prepared by at least one purification step.

[0085] An "antigenic region" or "antigenic determinant" or an "epitope" includes any protein determinant capable of specific binding to an antibody. This is the site on an antigen to which each distinct antibody molecule binds. Epitopic determinants usually consist of active surface groupings of molecules such as amino acids or sugar side chains and usually have specific three-dimensional structural characteristics, as well as charge characteristics.

[0086] "Antibody specificity" refers to an antibody that has a stronger binding affinity for an antigen from a first subject species than it has for a homologue of that antigen from a second subject species. Normally, the antibody "binds specifically" to a human antigen (i.e., has a binding affinity (Kd) value of no more than about 1×10^{-7} M, preferably no more than about 1×10^{-8} M and most preferably no more than about 1×10^{-9} M) but has a binding affinity for a homologue of the antigen from a second subject species which is at least about 50 fold, or at least about 500 fold, or at least about 1000 fold, weaker than its binding affinity for the human antigen. The antibody can be of any of the various types of antibodies as defined above, but preferably is a humanized or human antibody (see, e.g., Queen et al., U.S. Patent Nos. 5,530,101; 5,585,089; 5,693,762; and 6,180,370).

[0087] The present invention provides an "antibody variant," which refers to an amino acid sequence variant of an antibody wherein one or more of the amino acid residues have been modified. Such variants necessarily have less than 100% sequence identity or similarity with the amino acid sequence and have at least 75% amino acid sequence identity or similarity with the amino acid sequence of either the heavy or light chain variable domain of the antibody, more preferably at least 80%, more preferably at least 85%, more preferably at least 90%, and most preferably at least 95%. Since the method of the

invention applies equally to both polypeptides and antibodies and fragments thereof, these terms are sometimes employed interchangeably.

[0088] The term "antibody fragment" refers to a portion of a full-length antibody, including the antigen binding or variable region or the antigen-binding portion thereof. Examples of antibody fragments include Fab, Fab', $F(ab')_2$ and Fv fragments. Papain digestion of antibodies produces two identical antigen binding fragments, called the Fab fragment, each with a single antigen binding site, and a residual "Fc" fragment, so-called for its ability to crystallize readily. Pepsin treatment yields an $F(ab')_2$ fragment that has two antigen binding fragments which are capable of crosslinking antigen, and a residual other fragment (which is termed pFc'). Additional antigen-binding fragments can include diabodies, triabodies, tetrabodies, single-chain Fv, single-chain Fv-Fc, a SMIP, and multispecific antibodies formed from antibody fragments. As used herein, a "functional fragment" with respect to antibodies, refers to an Fv, F(ab), $F(ab')_2$ or other antigen-binding fragments comprising one or more CDRs that has the same antigen-binding specificity as an antibody.

[0089] An "Fv" fragment is the minimum antibody fragment that contains a complete antigen recognition and binding site. This region consists of a dimer of one heavy and one light chain variable domain in a tight, non-covalent association (V_H -V_L dimer). It is in this configuration that the three complementarity determining regions ("CDRs") of each variable domain interact to define an antigen-binding site on the surface of the V_H -V_L dimer. Collectively, the six CDRs confer antigen-binding specificity to the antibody. However, even a single variable domain (or half of an Fv comprising only three CDRs specific for an antigen) has the ability to recognize and bind antigen, although typically at a lower affinity than the entire binding site.

[0090] The Fab fragment [also designated as F(ab)] also contains the constant domain of the light chain and the first constant domain (CH1) of the heavy chain. Fab' fragments differ from Fab fragments by the addition of a few residues at the carboxyl terminus of the heavy chain CH1 domain including one or more cysteines from the antibody hinge region. Fab'-SH is the designation herein for Fab' in which the cysteine residue(s) of the constant

domains have a free thiol group. F(ab') fragments are produced by cleavage of the disulfide bond at the hinge cysteines of the $F(ab')_2$ pepsin digestion product. Additional chemical couplings of antibody fragments are known to those of ordinary skill in the art.

[0091] A "single-chain Fv" or "scFv" antibody fragment contains V_H and V_L domains, wherein these domains are present in a single polypeptide chain. Typically, the Fv polypeptide further comprises a polypeptide linker between the V_H and V_L domains which enables the scFv to form the desired structure for antigen binding. For a review of scFv, see Plückthun in The Pharmacology of Monoclonal Antibodies, vol. 113, Rosenburg and Moore eds., Springer-Verlag, New York, pp. 269-315 (1994). A single chain Fv-Fc is an scFv linked to a Fc region.

[0092] A "diabody" is a small antibody fragment with two antigen-binding sites, which fragments comprise a variable heavy domain (V_H) connected to a variable light domain (V_L) in the same polypeptide chain. By using a linker that is too short to allow pairing between the two domains on the same chain, the domains are forced to pair with the complementary domains of another chain and create two antigen-binding sites. Diabodies are described more fully in, for example, EP 0 404 097; WO 93/11161; and Hollinger et al., Proc. Natl. Acad. Sci. USA 90:6444-6448 (1993). Triabodies, tetrabodies and other antigen-binding antibody fragments have been described by Hollinger and Hudson, 2005, Nature Biotechnology 23:1126.

[0093] A "small modular immunopharmaceutical" or "SMIP" is a single-chain polypeptide including a binding domain (i.e., an scFv or an antigen binding portion of na antibody), a hinge region and an effector domain (e.g., an antibody Fc region or a portion thereof). SMIPs are described in Published U.S. Patent Application No. 2005-0238646.

[0094] The present invention further provides monoclonal antibodies, polyclonal antibodies as well as chimeric and humanized antibodies, and antigen-binding fragments thereof to prominin-1. In general, to generate antibodies, an isolated peptide is used as an immunogen and is administered to a mammalian organism, such as a rat, rabbit or mouse. The full-length prominin-1 protein, or an antigenic peptide fragment or a fusion protein thereof, can be used as an immunogen. Particularly important fragments are those covering

functional domains, such as the extracellular domain or a portion thereof. Many methods are known for generating and/or identifying antibodies to a given target peptide. Several such methods are described by Harlow, Antibodies, Cold Spring Harbor Press (1989); Harlow and Lane, Using Antibodies, Cold Spring Harbor Press (1998); Lane, R.D., 1985, *J. Immunol. Meth.* 81:223-228; Kubitz et al.., 1996, *J. Indust. Microbiol. Biotech.* 19:71-76; and Berry et al., 2003, Hybridoma and Hybridomics 22 (1): 23-31.

[0095] The term "monoclonal antibody" as used herein refers to an antibody obtained from a population of substantially homogeneous antibodies, i.e., the individual antibodies comprising the population are identical except for possible naturally occurring mutations that may be present in minor amounts. Monoclonal antibodies are highly specific, being directed against a single antigenic site. Furthermore, in contrast to conventional (polyclonal) antibody preparations which typically include different antibodies directed against different determinants (epitopes), each monoclonal antibody is directed against a single determinant on the antigen. In addition to their specificity, the monoclonal antibodies are advantageous in that substantially homogenous antibodies can be produced by a hybridoma culture which is uncontaminated by other immunoglobulins or antibodies. The modifier "monoclonal" antibody indicates the character of the antibody as being obtained from a substantially homogeneous population of antibodies, and is not to be construed as requiring production of the antibody by any particular method. For example, the monoclonal antibodies to be used in accordance with the present invention may be made by the hybridoma method first described by Kohler and Milstein, *Nature* 256: 495-497 (1975) or may be made by recombinant methods, e.g., as described in U.S. Pat. No. 4,816,567. The monoclonal antibodies for use with the present invention may also be isolated from phage antibody libraries using the techniques described in Clackson et al., *Nature* 352: 624-628 (1991), as well as in Marks et al., *J. Mol. Biol.* 222: 581-597 (1991).

[0096] "Humanized" forms of non-human (e.g., murine or rabbit) antibodies are chimeric immunoglobulins, immunoglobulin chains or fragments thereof (such as Fv, Fab, Fab', $F(ab')_2$ or other antigen-binding subsequences of antibodies) which contain minimal sequence derived from non-human immunoglobulin. For the most part, humanized antibodies are human immunoglobulins (a recipient antibody) in which residues from a

complementary determining region (CDR) of the recipient are replaced by residues from a CDR of a non-human species (a donor antibody) such as mouse, rat or rabbit having the desired specificity, affinity and capacity. In some instances, Fv framework residues of the human immunoglobulin are replaced by corresponding non-human residues. Furthermore, a humanized antibody may comprise residues which are found neither in the recipient antibody nor in the imported CDR or framework region (FR) sequences. These modifications are made to further refine and optimize antibody performance. In general, the humanized antibody will comprise substantially all of at least one, and typically two, variable domains, in which all or substantially all of the CDRs correspond to those of a non-human immunoglobulin and all or substantially all of the FRs are those of a human immunoglobulin consensus sequence. The humanized antibody optimally also will comprise at least a portion of an immunoglobulin constant region (Fc), typically that of a human immunoglobulin. For further details, see: Jones et al., Nature 321:522-525 (1986); Reichmann et al., Nature 332:323-327 (1988) and Presta, Curr. Op. Struct. Biol. 2:593-596 (1992).

[0097] Polyclonal antibodies may be prepared by any known method or modifications of these methods including obtaining antibodies from patients. For example, a complex of an immunogen such as prominin-1 protein, peptides or fragments thereof, and a carrier protein is prepared and an animal is immunized by the complex according to the same manner as that described with respect to the above monoclonal antibody preparation and the description in the Example. A serum or plasma containing the antibody against the protein is recovered from the immunized animal and the antibody is separated and purified. The gamma globulin fraction or the IgG antibodies can be obtained, for example, by use of saturated ammonium sulfate or DEAE SEPHADEX, or other techniques known to those skilled in the art.

[0098] The antibody titer in the antiserum can be measured according to the same manner as that described above with respect to the supernatant of the hybridoma culture. Separation and purification of the antibody can be carried out according to the same separation and purification method of antibody as that described with respect to the above monoclonal antibody and in the Example.

[0099] The protein used herein as the immunogen is not limited to any particular type of immunogen. In one aspect, antibodies are preferably prepared from regions or discrete fragments of the prominin-1 protein. Antibodies can be prepared from any region of the peptide as described herein. In particular, they are selected from a group consisting of SEQ ID NOS:1-5 and fragments thereof. An antigenic fragment will typically comprise at least 8 contiguous amino acid residues. The antigenic peptide can comprise, however, at least 10, 12, 14, 16 or more amino acid residues. Such fragments can be selected on a physical property, such as fragments that correspond to regions located on the surface of the protein, e.g., hydrophilic regions or that can be selected based on sequence uniqueness.

[00100] Antibodies may also be produced by inducing production in the lymphocyte population or by screening antibody libraries or panels of highly specific binding reagents as disclosed in Orlandi et al. (Proc. Natl. Acad. Sci. 86:3833-3837 (1989)) or Winter et al. (Nature 349:293-299 (1991)). A protein may be used in screening assays of phagemid or B-lymphocyte immunoglobulin libraries to identify antibodies having a desired specificity. Numerous protocols for competitive binding or immunoassays using either polyclonal or monoclonal antibodies with established specificities are well known in the art. See, e.g., Smith, Curr. Opin. Biotechnol. 2: 668-673 (1991).

[00101] The antibodies of the present invention can also be generated using various phage display methods known in the art. In phage display methods, functional antibody domains are displayed on the surface of phage particles which carry the polynucleotide sequences encoding them. In particular, such phage can be utilized to display antigen-binding domains expressed from a repertoire or combinatorial antibody library (e.g., human or murine). Phage expressing an antigen binding domain that binds the antigen of interest can be selected or identified with antigen, e.g., using labeled antigen or antigen bound or captured to a solid surface or bead. Phage used in these methods are typically filamentous phage including fd and M13 binding domains expressed from phage with Fab, Fv or disulfide stabilized Fv antibody domains recombinantly fused to either the phage gene III or gene VIII protein. Examples of phage display methods that can be used to make the antibodies of the present invention include those disclosed in Brinkman et al., J. Immunol. Methods 182:41-50 (1995); Ames et al., J. Immunol. Methods 184:177-186 (1995); Kettleborough et al., Eur. J. Immunol. 24:952-

958 (1994); Persic et al., Gene 187:9-18 (1997); Burton et al., Advances in Immunology 57:191-280 (1994); PCT application No. PCT/GB91/01134; PCT publications WO 90/02809; WO 91/10737; WO 92/01047; WO 92/18619; WO 93/11236; WO 95/15982; WO 95/20401; and U.S. Pat. Nos. 5,698,426; 5,223,409; 5,403,484; 5,580,717; 5,427,908; 5,750,753; 5,821,047; 5,571,698; 5,427,908; 5,516,637; 5,780,225; 5,658,727; 5,733,743 and 5,969,108; each of which is incorporated herein by reference in its entirety.

[00102] Antibodies, e.g., antibody variants, can be also made recombinantly. When using recombinant techniques, the antibody variant can be produced intracellularly, in the periplasmic space, or directly secreted into the medium. If the antibody variant is produced intracellularly, as a first step, the particulate debris, either host cells or lysed fragments, is removed, for example, by centrifugation or ultrafiltration. Carter et al., Bio/Technology 10: 163-167 (1992) describe a procedure for isolating antibodies that are secreted to the periplasmic space of *E. coli*. Briefly, cell paste is thawed in the presence of sodium acetate (pH 3.5), EDTA, and phenylmethylsulfonylfluoride (PMSF) over about 30 minutes. Cell debris can be removed by centrifugation. Where the antibody variant is secreted into the medium, supernatants from such expression systems are generally first concentrated using a commercially available protein concentration filter, for example, an Amicon or Millipore PELLICON ultrafiltration unit. A protease inhibitor such as PMSF may be included in any of the foregoing steps to inhibit proteolysis and antibiotics may be included to prevent the growth of adventitious contaminants.

[00103] The antibodies or antigen binding fragments may also be produced by genetic engineering. The technology for expression of both heavy and light chain genes in *E. coli* is the subject of PCT publication numbers WO 901443, WO901443, and WO 9014424 and in Huse et al., 1989 Science 246:1275-1281. The general recombinant methods are well known in the art.

[00104] The antibody composition prepared from the cells can be purified using, for example, hydroxylapatite chromatography, gel electrophoresis, dialysis, and/or affinity chromatography, with affinity chromatography being the preferred purification technique. The suitability of protein A as an affinity ligand depends on the species and isotype of the

immunoglobulin Fc domain of the antibody. Protein A can be used to purify antibodies that are based on human delta1, delta2 or delta4 heavy chains (Lindmark et al., J. Immunol Meth. 62: 1-13 (1983)). Protein G is recommended for all mouse isotypes and for human delta3 (Guss et al., EMBO J. 5: 1567-1575 (1986)). The matrix to which the affinity ligand is attached is most often agarose, but other matrices are available. Mechanically stable matrices such as controlled pore glass or poly(styrenedivinyl)benzene allow for faster flow rates and shorter processing times than can be achieved with agarose. Where the antibody comprises a CH3 domain, the BAKERBOND ABX™ resin (J.T. Baker, Phillipsburg, N.J.) is useful for purification. Other techniques for protein purification such as fractionation on an ion-exchange column, ethanol precipitation, reverse phase HPLC, chromatography on silica, chromatography on heparin hepharos, chromatography on an anion or cation exchange resin (such as a polyaspartic acid column), chromatofocusing, SDS-PAGE, and ammonium sulfate precipitation are also available, depending on the antibody to be recovered.

[00105] Following any preliminary purification step(s), contaminants in the mixture containing the antibody of interest may be removed by low pH hydrophobic interaction chromatography using an elution buffer at a pH between about 2.5-4.5, preferably performed at low salt concentrations (e.g., from about 0-0.25M salt).

[00106] A prominin-1 antibody also can be obtained from commercial sources. For example, antibodies CD133/1 (AC133) and CD133/2 (293C3) from Miltenyi Biotech. Antibody AC133 (32AT1672) can be obtained from Abgent). These and other sources of antibody can be used according to the present invention.

3. Antibody Drug Conjugates Against Prominin-1 Protein or Fragments Thereof

[00107] An antibody against Promin-1 may be coupled (e.g., covalently bonded) to a suitable therapeutic agent (as further discussed herein) either directly or indirectly (e.g., via a linker group). A direct reaction between an antibody and a therapeutic agent is possible when each possesses a substituent capable of reacting with the other. For example, a nucleophilic group, such as an amino or sulfhydryl group, on one molecule may be capable

of reacting with a carbonyl-containing group, such as an anhydride or an acid halide, or with an alkyl group containing a good leaving group (e.g., a halide) on the other molecule.

[00108] Alternatively, it may be desirable to couple a therapeutic agent and an antibody via a linker group. A linker group can function as a spacer to distance an antibody from an agent in order to avoid interference with binding capabilities. A linker group can also serve to increase the chemical reactivity of a substituent on an agent or an antibody, and thus increase the coupling efficiency. An increase in chemical reactivity may also facilitate the use of agents, or functional groups on agents, which otherwise would not be possible.

[00109] It will be evident to those skilled in the art that a variety of bifunctional or polyfunctional reagents, both homo- and hetero-functional (such as those described in the catalog of the Pierce Chemical Co., Rockford, Ill.), may be employed as the linker group. Coupling may be effected, for example, through amino groups, carboxyl groups, sulfhydryl groups or oxidized carbohydrate residues. There are numerous references describing such methodology, e.g. U.S. Pat. No. 4,671,958, to Rodwell et al.

[00110] Where a therapeutic agent is more potent when free from the antibody portion of the immunoconjugates of the present invention, it may be desirable to use a linker group which is cleavable during or upon internalization into a cell. A number of different cleavable linker groups have been described. The mechanisms for the intracellular release of an agent from these linker groups include cleavage by reduction of a disulfide bond (e.g., U.S. Pat. No. 4,489,710, to Spitler), by irradiation of a photolabile bond (e.g., U.S. Pat. No. 4,625,014, to Senter et al.), by hydrolysis of derivatized amino acid side chains (e.g., U.S. Pat. No. 4,638,045, to Kohn et al.), by serum complement-mediated hydrolysis (e.g., U.S. Pat. No. 4,671,958, to Rodwell et al.), by protease cleavable linker (e.g., U.S. Pat. No. 6,214,345), and acid-catalyzed hydrolysis (e.g., U.S. Pat. No. 4,569,789, to Blattler et al.).

[00111] It may be desirable to couple more than one agent to an antibody. In one embodiment, multiple molecules of an agent are coupled to one antibody molecule. In another embodiment, more than one type of agent may be coupled to one antibody.

Regardless of the particular embodiment, conjugates with more than one agent may be prepared in a variety of ways as described herein.

[00112] In some embodiments, an Antibody-Drug Conjugate has the following formula I:

[00113] $$Ab \left(A_a - W_w - Y_y - D \right)_p \qquad \text{I}$$

[00114] or a pharmaceutically acceptable salt or solvate thereof, wherein:

[00115] Ab- is an Antibody Unit;

[00116] -Aa-Ww-Yy- is a Linker Unit (LU), wherein:

[00117] -A- is a Stretcher Unit,

[00118] a is 0 or 1,

[00119] each -W- is independently an Amino Acid Unit,

[00120] w is an integer ranging from 0 to 12,

[00121] -Y- is a Spacer Unit, and

[00122] y is 0, 1 or 2;

[00123] p ranges from 1 to about 20; and

[00124] -D is a Drug Unit.

[00125] In some embodiments, an Antibody-Drug Conjugate has the following formula I:

[00126] $$Ab \left(D \right)_p \qquad \text{II}$$

[00127] or a pharmaceutically acceptable salt or solvate thereof, wherein:

[00128] Ab- is an Antibody Unit;

[00129] -D is a Drug Unit; and

[00130] p ranges from 1 to about 20.

[00131] The Antibody Unit comprises an antibody or antigen-binding antibody fragment. p, the number of Drug Units or Drug Linker Units attached to an Antibody Unit can range from about 1 to about 20 Drug Units per Antibody Unit, from about 1 to about 8 Drug Units per Antibody Unit, from about 2 to about 8 Drug Units per Antibody Unit, from about 2 to about 6, or from about 2 to about 4 Drug Units per Antibody Unit. In some embodiments, p is about 2, about 4, about 6 or about 8 Drug Units per Antibody Unit.

[00132] The average number of Drug Units per Antibody Unit in a preparation of conjugation reactions may be characterized by conventional means such as mass spectroscopy, ELISA, and HPLC. The quantitative distribution of Antibody-Drug-Conjugates in terms of p may also be determined. In some instances, separation, purification, and characterization of homogeneous Antibody-Drug Conjugates, where p is a certain value, from Antibody-Drug Conjugates with other drug loadings may be achieved by means such as reverse phase HPLC or electrophoresis (see, e.g., Hamblett et al., Clinical Cancer Res. 10:7063-70 (2004).

[00133] The Antibody Unit

[00134] An Antibody Unit can form a bond to a Stretcher Unit, an Amino Acid Unit, a Spacer Unit, or a Drug Unit, as more fully described infra. An Antibody Unit can form a bond to a Linker Unit via a heteroatom of the Antibody Unit. Heteroatoms that may be present on an Antibody Unit include sulfur (e.g., from a sulfhydryl group of an antibody), oxygen (e.g., from a carbonyl, carboxyl or hydroxyl group of an antibody) and nitrogen (e.g., from a primary or secondary amino group of an antibody). These heteroatoms can be present on the Antibody Unit in the unit's natural state, for example a naturally-occurring antibody, or can be introduced into the Antibody Unit via chemical modification.

[00135] In one embodiment, an Antibody Unit has a sulfhydryl group and the Antibody Unit bonds to the Linker Unit via the sulfhydryl group's sulfur atom.

[00136] In another embodiment, the Antibody Unit has a lysine residue(s) that can react with activated esters (such esters include, but are not limited to, N-hydroxysuccinimde, pentafluorophenyl, and p-nitrophenyl esters) of the Linker Unit and thus form an amide bond consisting of the nitrogen atom of the Antibody Unit and the C=O group of the Linker Unit.

[00137] In yet another aspect, the Antibody Unit has one or more lysine residues that can be chemically modified to introduce one or more sulfhydryl groups. The Antibody Unit bonds to the Linker Unit via the sulfhydryl group's sulfur atom. Reagents that can be used to modify lysines include, but are not limited to, N-succinimidyl S-acetylthioacetate (SATA) and 2-Iminothiolane hydrochloride (Traut's Reagent).

[00138] In another embodiment, the Antibody Unit can have one or more carbohydrate groups that can be chemically modified to have one or more sulfhydryl groups. The Antibody Unit bonds to the Linker Unit, such as the Stretcher Unit, via the sulfhydryl group's sulfur atom.

[00139] In yet another embodiment, the Antibody Unit can have one or more carbohydrate groups that can be oxidized to provide an aldehyde (-CHO) group (see, e.g., Laguzza et al., J. Med. Chem. 32(3):548-55 (1989)). The corresponding aldehyde can form a bond with a reactive site on a Stretcher. Reactive sites on a Stretcher that can react with a carbonyl group on an Antibody Unit include, but are not limited to, hydrazine and hydroxylamine. Other protocols for the modification of proteins for the attachment or association of Drug Units are described in Coligan et al., Current Protocols in Protein Science, vol. 2, John Wiley & Sons (2002).

[00140] A cysteine residue also can be introduced into a protein using recombinant DNA technology. For example, a cysteine residue(s) can be introduced into a protein by mutagenesis of a nucleic acid encoding the protein. See generally Sambrook et al., Molecular Cloning, A Laboratory Manual, 3rd ed., Cold Spring Harbor Publish., Cold Spring Harbor, New York (2001); Ausubel et al., Current Protocols in Molecular Biology, 4th ed., John Wiley and Sons, New York (1999)) Sulfhydryl groups can be introduced into a protein, for example, within the polypeptide or at the carboxy-terminus.

Linker Units

1] A "Linker Unit" (LU) is a bifunctional compound which can be used to link a Drug and an Antibody Unit to form Antibody Drug conjugate compounds. In some embodiments, nker Unit has the formula:

$$-A_a-W_w-Y_y-$$

wherein: -A- is a Stretcher Unit,

a is 0 or 1,

each -W- is independently an Amino Acid Unit,

w is an integer ranging from 0 to 12,

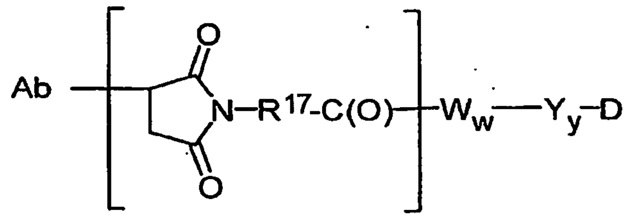
-Y- is a self-immolative Spacer Unit, and

y is 0, 1 or 2.

In some embodiments, a is 0 or 1, w is 0 or 1, and y is 0, 1 or 2. In some embodiments, a is 0 or 1, w is 0 or 1, and y is 0 or 1.

The Stretcher Unit

[00142] The Stretcher Unit (A), when present, is capable of linking an Antibody Unit to an Amino Acid Unit (-W-), if present, to a Spacer Unit (-Y-), if present; or to a Drug Unit (-D). The Stretcher Unit can form a bond with a sulfur atom of the Antibody Unit. The sulfur atom can be derived from a sulfhydryl group of an Antibody. Representative Stretcher Units of this embodiment are depicted within the square brackets of Formulas IIIa and IIIb, wherein L-, -W-, -Y-, -D, w and y are as defined above, and R_{17} is selected from $-C_1-C_{10}$ alkylene-, $-C_3-C_8$ carbocyclo-, $-O-(C_1-C_8$ alkyl)-, -arylene-, $-C_1-C_{10}$ alkylene-arylene-, -arylene-C_1-C_{10} alkylene-, $-C_1-C_{10}$ alkylene-(C_3-C_8 carbocyclo)-, -(C_3-C_8 carbocyclo)-C_1-C_{10} alkylene-, $-C_3-C_8$ heterocyclo-, $-C_1-C_{10}$ alkylene-(C_3-C_8 heterocyclo)-, -(C_3-C_8 heterocyclo)-C_1-C_{10} alkylene-, $-(CH_2CH_2O)_r-$, and $-(CH_2CH_2O)_r-CH_2-$; and r is an integer ranging from 1-10. It is to be understood from all the exemplary embodiments of Formula I, such as III-VI, that even where not denoted expressly, from 1 to 20 drug moieties are linked to an Antibody Unit (p = 1-20).

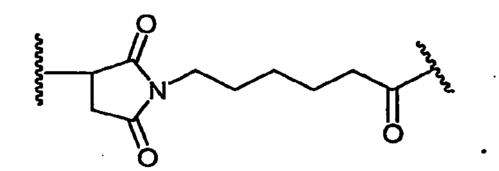


IIIa

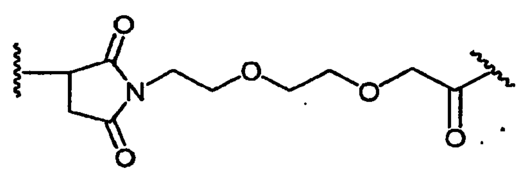
$$Ab \left[CH_2-CONH-R^{17}-C(O) \right] W_w-Y_y-D$$

IIIb

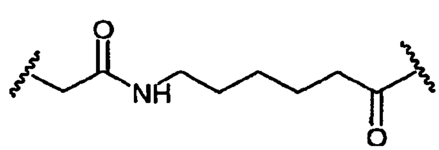
[00143] An illustrative Stretcher Unit is that of Formula IIIa wherein R^{17} is $-(CH_2)_5-$:



[00144] Another illustrative Stretcher Unit is that of Formula IIIa wherein R^{17} is $-(CH_2CH_2O)_r-CH_2-$; and r is 2:



[00145] Still another illustrative Stretcher Unit is that of Formula IIIb wherein R^{17} is $-(CH_2)_5-$:



[00146] The Stretcher Unit also can be linked to the Antibody Unit via a disulfide bond between a sulfur atom of the Antibody Unit and a sulfur atom of the Stretcher Unit. A

representative Stretcher Unit is depicted within the square brackets of Formula IV, wherein R^{17}, L-, -W-, -Y-, -D, w and y are as defined above.

$$\mathrm{Ab} - \left[\mathrm{S} - \mathrm{R}^{17} - \mathrm{C(O)} \right] - \mathrm{W}_w - \mathrm{Y}_y - \mathrm{D} \qquad \mathbf{IV}$$

[00147] In yet another embodiment, the reactive group of the Stretcher contains a reactive site that can form a bond with a primary or secondary amino group of an Antibody Unit. Examples of these reactive sites include, but are not limited to, activated esters such as succinimide esters, 4 nitrophenyl esters, pentafluorophenyl esters, tetrafluorophenyl esters, anhydrides, acid chlorides, sulfonyl chlorides, isocyanates and isothiocyanates. Representative Stretcher Units of this embodiment are depicted within the square brackets of Formulas Va and Vb, wherein $-R^{17}-$, L-, -W-, -Y-, -D, w and y are as defined above;

$$\mathrm{Ab} - \left[\mathrm{C(O)NH} - \mathrm{R}^{17} - \mathrm{C(O)} \right] - \mathrm{W}_w - \mathrm{Y}_y - \mathrm{D} \qquad \mathbf{Va}$$

$$\mathbf{Ab} - \left[\overset{\overset{\mathrm{S}}{\|}}{\mathrm{C}}\mathrm{NH} - \mathrm{R}^{17} - \mathrm{C(O)} \right] - \mathrm{W}_w - \mathrm{Y}_y - \mathrm{D} \qquad \mathbf{Vb}$$

[00148] In some embodiments, the reactive group of the Stretcher contains a reactive site that is reactive to a modified carbohydrate's (-CHO) group that can be present on an Antibody Unit. For example, a carbohydrate can be mildly oxidized using a reagent such as sodium periodate and the resulting (-CHO) Unit of the oxidized carbohydrate can be condensed with a Stretcher that contains a functionality such as a hydrazide, an oxime, a primary or secondary amine, a hydrazine, a thiosemicarbazone, a hydrazine carboxylate, and an arylhydrazide such as those described by Kaneko *et al.*, *Bioconjugate Chem.* 2:133 41 (1991). Representative Stretcher Units are depicted within the square brackets of

Formulas VIa, VIb, and VIc, wherein $-R_{17}-$, L-, -W-, -Y-, -D, w and y are as defined above.

$$\mathrm{Ab}\left[=\mathrm{N\text{-}NH}-\mathrm{R}^{17}\text{-}\mathrm{C(O)}\right]-\mathrm{W_w}-\mathrm{Y_y}-\mathrm{D}$$

VIa

$$\mathrm{Ab}\left[=\mathrm{N\text{-}O}-\mathrm{R}^{17}\text{-}\mathrm{C(O)}\right]-\mathrm{W_w}-\mathrm{Y_y}-\mathrm{D}$$

VIb

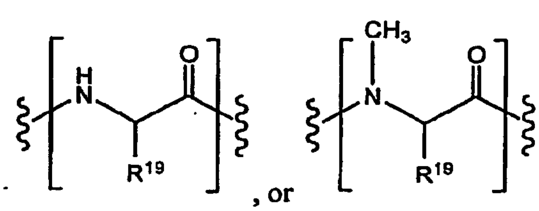
$$\mathrm{Ab}\left[=\mathrm{N\text{-}NH}-\overset{\displaystyle \mathrm{O}}{\overset{\|}{\mathrm{C}}}-\mathrm{R}^{17}\text{-}\mathrm{C(O)}\right]-\mathrm{W_w}-\mathrm{Y_y}-\mathrm{D}$$

VI

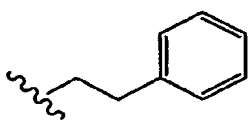
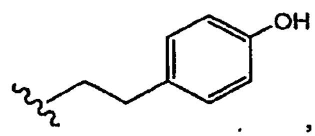
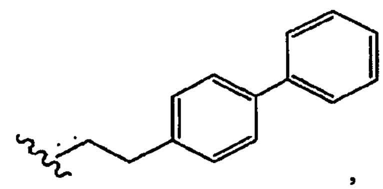
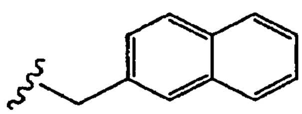
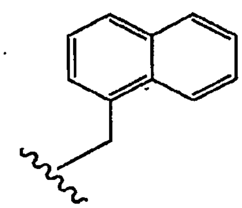
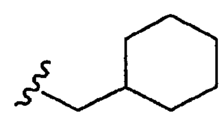
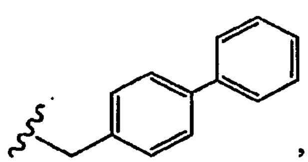
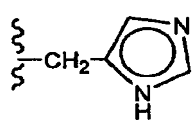
The Amino Acid Unit

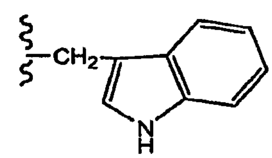
[00149] The Amino Acid Unit (-W-), when present, links the Stretcher Unit to the Spacer Unit if the Spacer Unit is present, links the Stretcher Unit to the Drug moiety if the Spacer Unit is absent, and links the Antibody Unit to the Drug Unit if the Stretcher Unit and Spacer Unit are absent.

[00150] W_w- is a dipeptide, tripeptide, tetrapeptide, pentapeptide, hexapeptide, heptapeptide, octapeptide, nonapeptide, decapeptide, undecapeptide or dodecapeptide unit. Each -W- unit independently has the formula denoted below in the square brackets, and w is an integer ranging from 0 to 12:



[00151] wherein R^{19} is hydrogen, methyl, isopropyl, isobutyl, *sec*-butyl, benzyl, *p*-hydroxybenzyl, $-CH_2OH$, $-CH(OH)CH_3$, $-CH_2CH_2SCH_3$, $-CH_2CONH_2$, $-CH_2COOH$, $-CH_2CH_2CONH_2$, $-CH_2CH_2COOH$, $-(CH_2)_3NHC(=NH)NH_2$, $-(CH_2)_3NH_2$, $-(CH_2)_3NHCOCH_3$, $-(CH_2)_3NHCHO$, $-(CH_2)_4NHC(=NH)NH_2$, $-(CH_2)_4NH_2$, $-(CH_2)_4NHCOCH_3$, $-(CH_2)_4NHCHO$, $-(CH_2)_3NHCONH_2$, $-(CH_2)_4NHCONH_2$, $-CH_2CH_2CH(OH)CH_2NH_2$, 2-pyridylmethyl-, 3-pyridylmethyl-, 4-pyridylmethyl-, phenyl, cyclohexyl,

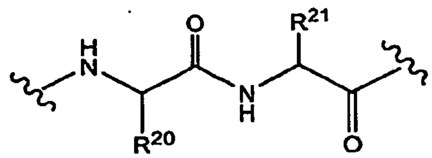

,

,

,

,

,

,

,


, or


.

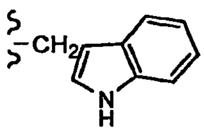
[00152] In some embodiments, the Amino Acid Unit can be enzymatically cleaved by one or more enzymes, including a cancer or tumor-associated protease, to liberate the Drug Unit (-D), which is protonated in vivo upon release to provide a Drug (D).

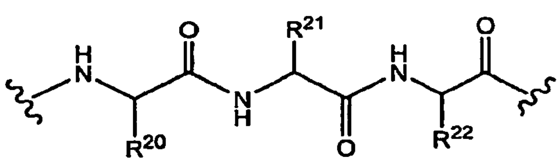
[00153] The Amino Acid Unit can comprise natural amino acids or non-natural amino acids. Illustrative Ww units are represented by formulas (VII)-(IX):



(VII)

wherein R^{20} and R^{21} are as follows:

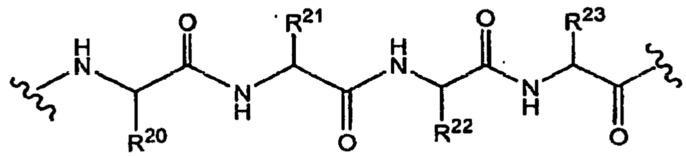
R^{20}	R^{21}
Benzyl	$(CH_2)_4NH_2$;
methyl	$(CH_2)_4NH_2$;
isopropyl	$(CH_2)_4NH_2$;
isopropyl	$(CH_2)_3NHCONH_2$;
benzyl	$(CH_2)_3NHCONH_2$;
isobutyl	$(CH_2)_3NHCONH_2$;
sec-butyl	$(CH_2)_3NHCONH_2$;
	$(CH_2)_3NHCONH_2$;
benzyl	methyl; and
benzyl	$(CH_2)_3NHC(=NH)NH_2$;



(VIII)

wherein R^{20}, R^{21} and R^{22} are as follows:

R^{20}	R^{21}	R^{22}
benzyl	benzyl	$(CH_2)_4NH_2$;
isopropyl	benzyl	$(CH_2)_4NH_2$; and
H	benzyl	$(CH_2)_4NH_2$;



(IX)

wherein R^{20}, R^{21}, R^{22} and R^{23} are as follows:

R^{20}	R^{21}	R^{22}	R^{23}
H	benzyl	isobutyl	H; and
methyl	isobutyl	methyl	isobutyl.

[00154] Exemplary Amino Acid Units include, but are not limited to, units of formula (VII) where: R^{20} is benzyl and R^{21} is $-(CH_2)_4NH_2$; R^{20} isopropyl and R^{21} is $-(CH_2)_4NH_2$; R^{20} isopropyl and R^{21} is $-(CH_2)_3NHCONH_2$. Another exemplary Amino Acid Unit is a unit of formula (VIII) wherein R^{20} is benzyl, R^{21} is benzyl, and R^{22} is $-(CH_2)_4NH_2$.

[00155] Useful $-W_w-$ units can be designed and optimized in their selectivity for enzymatic cleavage by a particular enzyme, for example, a tumor-associated protease. In one embodiment, a $-W_w-$ unit is that whose cleavage is catalyzed by cathepsin B, C and D, or a plasmin protease.

[00156] In one embodiment, $-W_w-$ is a dipeptide, tripeptide, tetrapeptide or pentapeptide. When R^{19}, R^{20}, R^{21}, R^{22} or R^{23} is other than hydrogen, the carbon atom to which R^{19}, R^{20}, R^{21}, R^{22} or R^{23} is attached is chiral.

[00157] Each carbon atom to which R^{19}, R^{20}, R^{21}, R^{22} or R^{23} is attached is independently in the (S) or (R) configuration.

[00158] The Amino Acid Unit can be, for example, valine-citrulline (vc), phenylalanine-lysine (k), N-methylvaline-citrulline, 5-aminovaleric acid, homo phenylalanine lysine, tetraisoquinolinecarboxylate lysine, cyclohexylalanine lysine, isonepecotic acid lysine, beta-alanine lysine, glycine serine valine glutamine or isonepecotic acid.

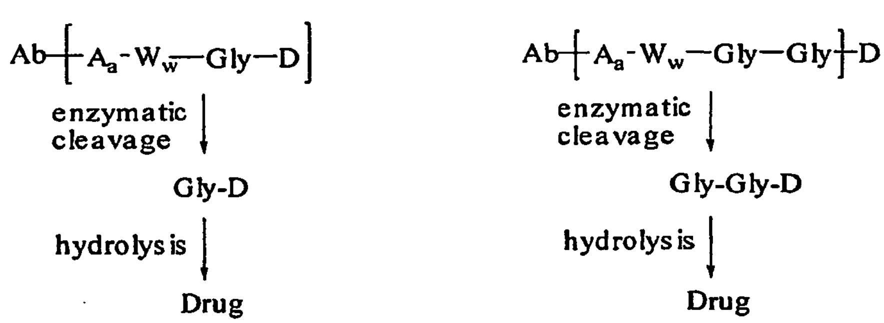
<u>The Spacer Unit</u>

[00159] The Spacer Unit (-Y-), when present, links an Amino Acid Unit to the Drug Unit when an Amino Acid Unit is present. Alternately, the Spacer Unit links the Stretcher Unit to the Drug Unit when the Amino Acid Unit is absent. The Spacer Unit also links the Drug Unit to the Antibody Unit when both the Amino Acid Unit and Stretcher Unit are absent.

[00160] Spacer Units are of two general types: non self-immolative or self-immolative. A non self-immolative Spacer Unit is one in which part or all of the Spacer Unit remains bound to the Drug moiety after cleavage, particularly enzymatic, of an Amino Acid Unit from the Antibody Drug Conjugate. Examples of a non self-immolative Spacer Unit include, but are not limited to a (glycine-glycine) Spacer Unit and a glycine Spacer Unit (both depicted in Scheme 1) (infra). When a conjugate containing a glycine-glycine Spacer Unit or a glycine Spacer Unit undergoes enzymatic cleavage via an enzyme (e.g., a tumor-cell associated-protease, a cancer-cell-associated protease or a lymphocyte-associated protease), a glycine-glycine-Drug moiety or a glycine-Drug moiety is cleaved from L-Aa-Ww-. In one embodiment, an independent hydrolysis reaction takes place within the target cell, cleaving the glycine-Drug moiety bond and liberating the Drug.

[00161] In another embodiment, $-Y_y-$ is a p-aminobenzyl alcohol (PAB) unit (see Schemes 2 and 3) whose phenylene portion is substituted with Q_m wherein Q is $-C_1-C_8$ alkyl, $-O-(C_1-C_8$ alkyl), -halogen,- nitro or -cyano; and m is an integer ranging from 0-4.

Scheme 1



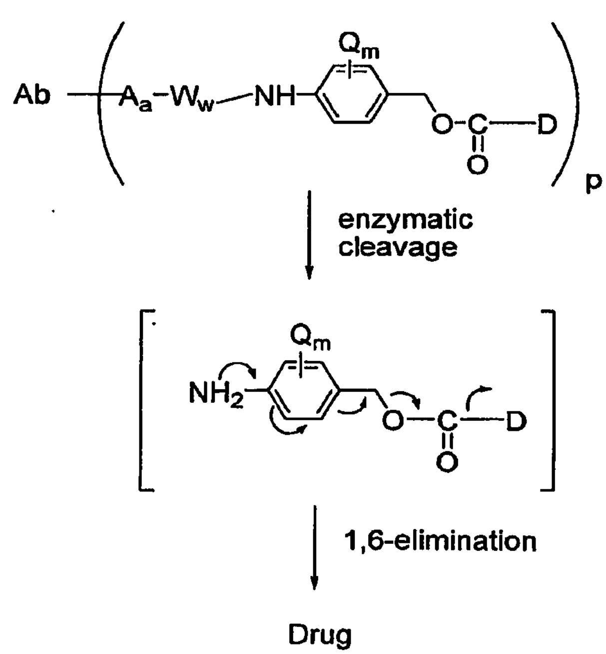
[00162] In some embodiments, a non self-immolative the Spacer Unit (-Y-) is -Gly-. In some embodiments, a non self-immolative Spacer Unit (-Y-) is -Gly-Gly-.

[00163] In one embodiment, an antibody drug conjugate compound is provided in which the Spacer Unit is absent (y=0), or a pharmaceutically acceptable salt or solvate thereof.

[00164] Alternatively, a conjugate containing a self-immolative Spacer Unit can release -D. As used herein, the term "self-immolative Spacer" refers to a bifunctional chemical moiety that is capable of covalently linking together two spaced chemical moieties into a stable tripartite molecule. It will spontaneously separate from the second chemical moiety if its bond to the first moiety is cleaved.

[00165] In some embodiments, -Y- is a PAB group that is linked to $-W_w-$ via the amino nitrogen atom of the PAB group, and connected directly to -D via a carbonate, carbamate or ether group. Without being bound by any particular theory or mechanism, Scheme 2 depicts a possible mechanism of Drug release of a PAB group which is attached directly to -D via a carbamate or carbonate group as described by Toki et al., *J. Org. Chem.* 67:1866-1872 (2002)).

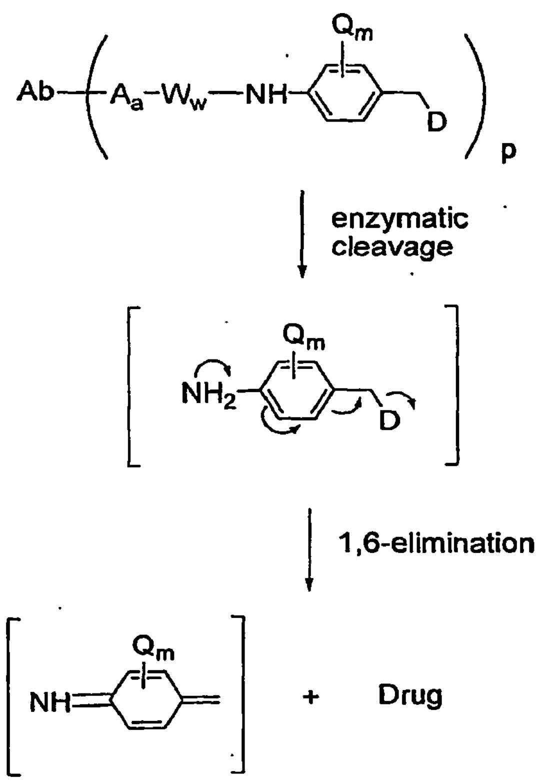
Scheme 2



[00166] In Scheme 2, Q is $-C_1-C_8$ alkyl, $-O-(C_1-C_8$ alkyl), -halogen, -nitro or -cyano; m is an integer ranging from 0-4; and p ranges from 1 to about 20.

[00167] Without being bound by any particular theory or mechanism, Scheme 3 depicts a possible mechanism of Drug release of a PAB group which is attached directly to -D via an ether or amine linkage, wherein D includes the oxygen or nitrogen group is part of the Drug Unit.

Scheme 3



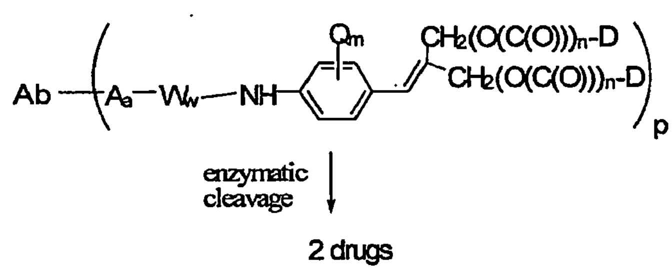
[00168] In Scheme 3, Q is $-C_1-C_8$ alkyl, $-O-(C_1-C_8$ alkyl), -halogen, -nitro or -cyano; m is an integer ranging from 0-4; and p ranges from 1 to about 20.

[00169] Other examples of self-immolative spacers include, but are not limited to, aromatic compounds that are electronically similar to the PAB group such as 2-

aminoimidazol-5-methanol derivatives (Hay et al., 1999, *Bioorg. Med. Chem. Lett.* 9:2237) and ortho or para-aminobenzylacetals. Spacers can be used that undergo cyclization upon amide bond hydrolysis, such as substituted and unsubstituted 4-aminobutyric acid amides (Rodrigues et al., *Chemistry Biology* 2:223 (1995)), appropriately substituted bicyclo[2.2.1] and bicyclo[2.2.2] ring systems (Storm et al., *J. Amer. Chem. Soc.* 94:5815 (1972)) and 2-aminophenylpropionic acid amides (Amsberry et al., *J. Org. Chem.* 55:5867 (1990)). Elimination of amine-containing drugs that are substituted at the α-position of glycine (Kingsbury et al., 1984, *J. Med. Chem.* 27:1447) are also examples of self-immolative spacers.

[00170] In one embodiment, the Spacer Unit is a branched bis(hydroxymethyl)-styrene (BHMS) unit as depicted in Scheme 4, which can be used to incorporate and release multiple drugs.

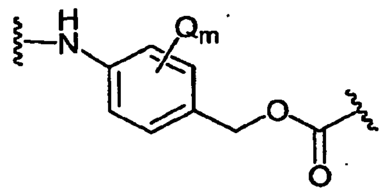
Scheme 4



[00171] In Scheme 4, Q is $-C_1-C_8$ alkyl, $-O-(C_1-C_8$ alkyl), -halogen, -nitro or -cyano; m is an integer ranging from 0-4; n is 0 or 1; and p ranges raging from 1 to about 20.

[00172] In some embodiments, the -D moieties are the same. In yet another embodiment, the -D moieties are different.

[00173] In one aspect, Spacer Units ($-Y_y-$) are represented by Formulas (X)-(XII):



X

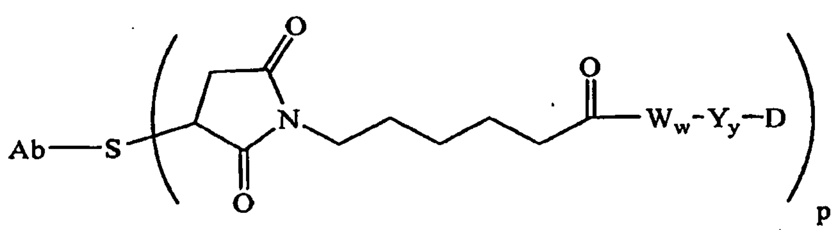
wherein Q is $-C_1-C_8$ alkyl, $-O-(C_1-C_8$ alkyl), -halogen, -nitro or -cyano; and m is an integer ranging from 0-4;

$$\text{-HN-CH}_2\text{-CO-} \quad \text{XI}$$

and

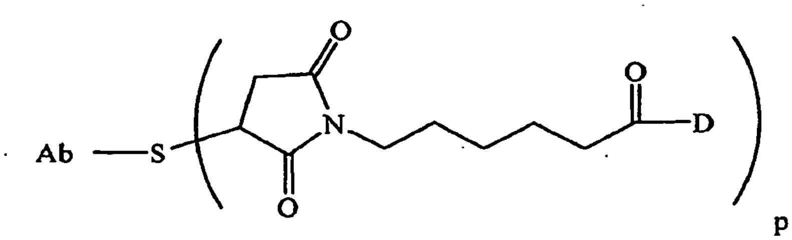
$$\text{-NHCH}_2\text{C(O)-NHCH}_2\text{C(O)-} \quad \text{XII.}$$

[00174] Embodiments of the Formula I, comprising an antibody-drug conjugate can include:

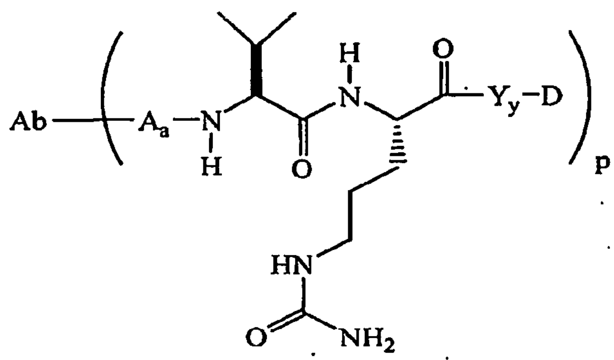


wherein w and y are each 0, 1 or 2,

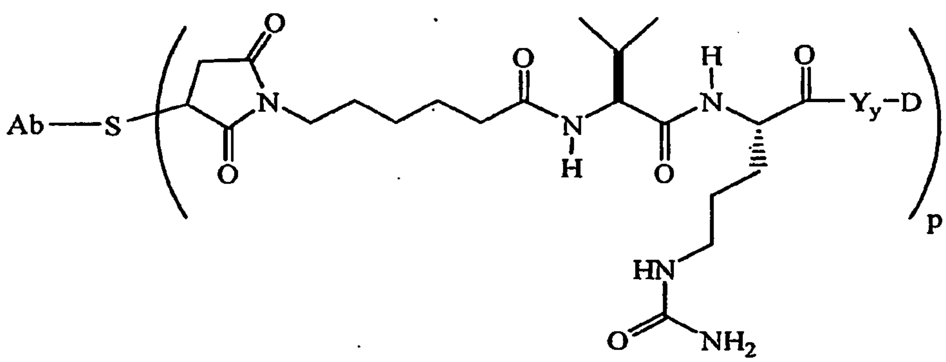
and,



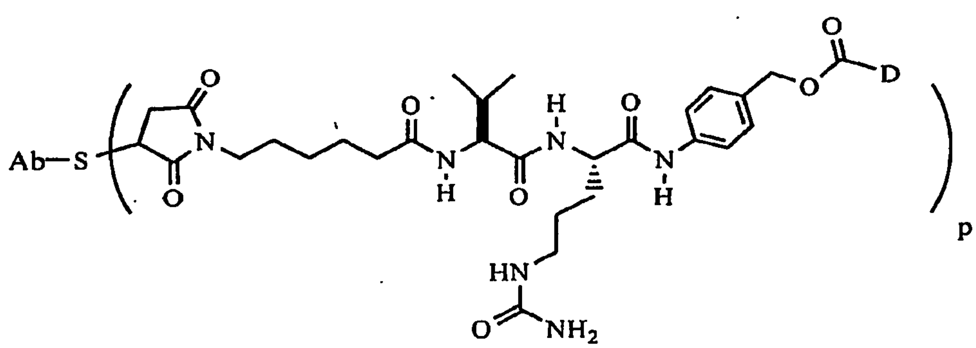
wherein w and y are each 0,



,



, and



.

The Drug Unit

[00175] The Drug Unit (D) can be any therapeutic agent. D has an atom that can form a bond with the Spacer Unit, with the Amino Acid Unit, with the Stretcher Unit or with the Antibody Unit. In some embodiments, the Drug Unit has a nitrogen atom that can form a bond with the Spacer Unit. As used herein, the terms "Drug Unit" and "drug moiety" are synonymous and used interchangeably.

[00176] The term "Drug Unit" as used herein refers to a therapeutic agent such as a chemotherapeutic agent (e.g., a cytotoxic or cytostatic agent or immunomodulatory agent), a radiotherapeutic agent, a therapeutic antibody, a small molecule (i.e., a chemical) drug, a peptide drug, an immunomodulatory agent, a differentiation inducer or a toxin, that is administered to a mammal, preferably a human, in need thereof.

[00177] Useful classes of cytotoxic or immunomodulatory agents include, for example, antitubulin agents, auristatins, DNA minor groove binders, DNA replication inhibitors, alkylating agents (e.g., platinum complexes such as cis-platin, mono(platinum), bis(platinum) and tri-nuclear platinum complexes and carboplatin), anthracyclines, antibiotics, antifolates, antimetabolites, chemotherapy sensitizers, duocarmycins, etoposides, fluorinated pyrimidines, ionophores, lexitropsins, nitrosoureas, platinols, pre-forming compounds, purine antimetabolites, puromycins, radiation sensitizers, steroids, taxanes, topoisomerase inhibitors, vinca alkaloids, or the like.

[00178] Individual cytotoxic or immunomodulatory agents include, for example, an androgen, anthramycin (AMC), asparaginase, 5-azacytidine, azathioprine, bleomycin, busulfan, buthionine sulfoximine, calicheamicin or a calicheamicin derivative, a camptothecin or a camptothecins derivative, carboplatin, carmustine (BSNU), CC-1065, chlorambucil, cisplatin, colchicine, cyclophosphamide, cytidine arabinoside (cytarabine), cytochalasin B, dacarbazine, dactinomycin (formerly actinomycin), daunorubicin, decarbazine, docetaxel, doxorubicin, etoposide, an estrogen, 5-fluordeoxyuridine, 5-fluorouracil, gemcitabine, gramicidin D, hydroxyurea, idarubicin, ifosfamide, irinotecan, lomustine (CCNU), maytansine, mechlorethamine, melphalan, 6-mercaptopurine, methotrexate, mithramycin, mitomycin C, mitoxantrone, nitroimidazole, paclitaxel,

palytoxin, plicamycin, procarbizine, rhizoxin, streptozotocin, tenoposide, 6-thioguanine, thioTEPA, topotecan, vinblastine, vincristine, vinorelbine, VP-16 and VM-26.

[00179] Other suitable cytotoxic agents include, for example, DNA minor groove binders (e.g., enediynes and lexitropsins, a CBI compound; see also U.S. Patent No. 6,130,237), duocarmycins, taxanes (e.g., paclitaxel and docetaxel), puromycins, vinca alkaloids, CC-1065, SN-38, topotecan, morpholino-doxorubicin, rhizoxin, cyanomorpholino-doxorubicin, echinomycin, combretastatin, netropsin, epothilone A and B, estramustine, cryptophysins, cemadotin, a maytansinoid, discodermolide, eleutherobin, and mitoxantrone.

[00180] Other suitable agents include radionuclides, differentiation inducers, drugs, toxins, and derivatives thereof. Preferred radionuclides include ^{90}Y, ^{123}I, ^{125}I, ^{131}I, ^{186}Re, ^{188}Re ^{211}At, and ^{212}Bi. Preferred drugs include methotrexate, and pyrimidine and purine analogs. Preferred differentiation inducers include phorbol esters and butyric acid. Preferred toxins include ricin, abrin, diptheria toxin, cholera toxin, gelonin, Pseudomonas exotoxin, Shigella toxin, and pokeweed antiviral protein.

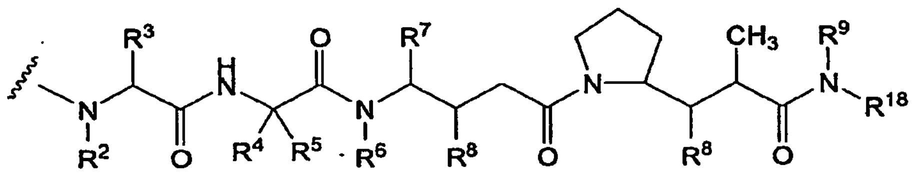
[00181] In some embodiments, the Drug Unit is an anti-tubulin agent. Examples of anti-tubulin agents include, but are not limited to, taxanes (e.g., Taxol® (paclitaxel), Taxotere® (docetaxel)), T67 (Tularik) and vinca alkyloids (e.g., vincristine, vinblastine, vindesine, and vinorelbine). Other antitubulin agents include, for example, baccatin derivatives, taxane analogs (e.g., epothilone A and B), nocodazole, colchicine and colcimid, estramustine, cryptophysins, cemadotin, a maytansinoid, combretastatins, discodermolide, and eleutherobin.

[00182] In certain embodiments, the cytotoxic agent is a maytansinoid, another group of anti-tubulin agents. For example, in specific embodiments, the maytansinoid is maytansine, DM-1 (ImmunoGen, Inc.; see also Chari et al., Cancer Res. 52:127-131 (1992)) or DM-4.

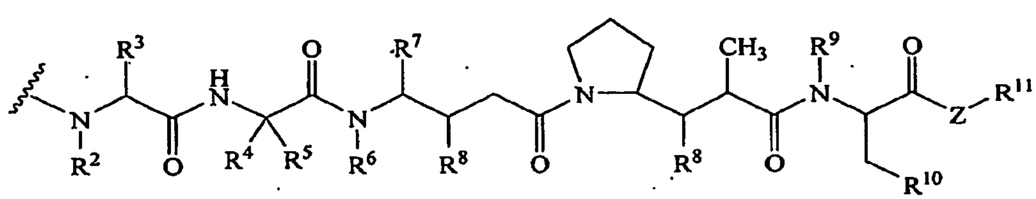
[00183] In some embodiments, the Drug is an auristatin, such as auristatin E (also known in the art as dolastatin-10) or a derivative thereof. Typically, the auristatin E derivative is, e.g., an ester formed between auristatin E and a keto acid. For example, auristatin E can be

reacted with paraacetyl benzoic acid or benzoylvaleric acid to produce AEB and AEVB, respectively. Other typical auristatin derivatives include AFP, MMAF, and MMAE. The synthesis and structure of auristatin derivatives are described in U.S. Patent Application Publication Nos. 2003-0083263, 2005-0238649 and 2005-0009751; PCT Publication Nos WO 04/010957 and WO 02/088172, and U.S. Patent Nos. 6,323,315; 6,239,104; 6,034,065; 5,780,588; 5,665,860; 5,663,149; 5,635,483; 5,599,902; 5,554,725; 5,530,097; 5,521,284; 5,504,191; 5,410,024; 5,138,036; 5,076,973; 4,986,988; 4,978,744; 4,879,278; 4,816,444; and 4,486,414.

[00184] In some embodiments, -D is either formula D_E or D_F:



D_E



D_F

wherein, independently at each location:

R^2 is selected from H and C_1-C_8 alkyl;

R^3 is selected from H, C_1-C_8 alkyl, C_3-C_8 carbocycle, aryl, C_1-C_{10} alkyl-aryl, C_1-C_{10} alkyl-(C_3-C_8 carbocycle), C_3-C_8 heterocycle and C_1-C_{10} alkyl-(C_3-C_8 heterocycle);

R^4 is selected from H, C_1-C_8 alkyl, C_3-C_8 carbocycle, aryl, C_1-C_{10} alkyl-aryl, C_1-C_{10} alkyl-(C_3-C_8 carbocycle), C_3-C_8 heterocycle and C_1-C_{10} alkyl-(C_3-C_8 heterocycle);

R^5 is selected from H and methyl;

or R^4 and R^5 jointly form a carbocyclic ring and have the formula $-(CR^aR^b)_n-$ wherein R^a and R^b are independently selected from H, C_1-C_8 alkyl and C_3-C_8 carbocycle and n is selected from 2, 3, 4, 5 and 6;

R^6 is selected from H and C_1-C_8 alkyl;

R^7 is selected from H, C_1-C_8 alkyl, C_3-C_8 carbocycle, aryl, C_1-C_{10} alkyl-aryl, C_1-C_{10} alkyl-(C_3-C_8 carbocycle), C_3-C_8 heterocycle and C_1-C_{10} alkyl-(C_3-C_8 heterocycle);
each R^8 is independently selected from H, OH, C_1-C_8 alkyl, C_3-C_8 carbocycle and O-(C_1-C_8 alkyl);
R^9 is selected from H and C_1-C_8 alkyl;
R^{10} is selected from aryl or C_3-C_8 heterocycle;
Z is O, S, NH, or NR^{12}, wherein R^{12} is C_1-C_8 alkyl;
R^{11} is selected from H, C_1-C_{20} alkyl, aryl, C_3-C_8 heterocycle, -$(R^{13}O)_m$-R^{14}, or -$(R^{13}O)_m$-$CH(R^{15})_2$;
m is an integer ranging from 1-1000;
R^{13} is C_2-C_8 alkyl;
R^{14} is H or C_1-C_8 alkyl;
each occurrence of R^{15} is independently H, COOH, $-(CH_2)_n-N(R^{16})_2$, $-(CH_2)_n-SO_3H$, or $-(CH_2)_n-SO_3-C_1-C_8$ alkyl;
each occurrence of R^{16} is independently H, C_1-C_8 alkyl, or $-(CH_2)_n-COOH$;
R^{18} is selected from $-C(R^8)_2-C(R^8)_2-$aryl, $-C(R^8)_2-C(R^8)_2-(C_3$-C_8 heterocycle), and $-C(R^8)_2-C(R^8)_2-(C_3$-C_8 carbocycle); and
n is an integer ranging from 0 to 6;

or a pharmaceutically acceptable salt or solvate thereof.

[00185] In one embodiment, R^3, R^4 and R^7 are independently isopropyl or sec-butyl and R^5 is -H. In an exemplary embodiment, R^3 and R^4 are each isopropyl, R^5 is H, and R^7 is sec-butyl.

[00186] In another embodiment, R^2 and R^6 are each methyl, and R^9 is H.

[00187] In still another embodiment, each occurrence of R^8 is $-OCH_3$.

[00188] In an exemplary embodiment, R^3 and R^4 are each isopropyl, R^2 and R^6 are each methyl, R^5 is H, R^7 is sec-butyl, each occurrence of R^8 is $-OCH_3$, and R^9 is H.

[00189] In one embodiment, Z is -O- or -NH-.

[00190] In one embodiment, R^{10} is aryl.

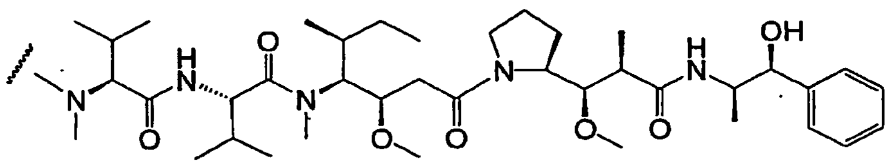
[00191] In an exemplary embodiment, R^{10} is -phenyl.

[00192] In an exemplary embodiment, when Z is -O-, R^{11} is H, methyl or t-butyl.

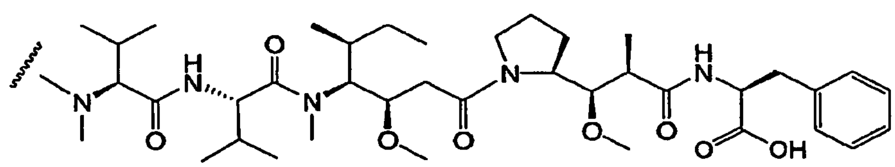
[00193] In one embodiment, when Z is -NH, R^{11} is $-CH(R^{15})_2$, wherein R^{15} is $-(CH_2)_n-N(R^{16})_2$, and R^{16} is $-C_1-C_8$ alkyl or $-(CH_2)_n-COOH$.

[00194] In another embodiment, when Z is -NH, R^{11} is $-CH(R^{15})_2$, wherein R^{15} is $-(CH_2)_n-SO_3H$.

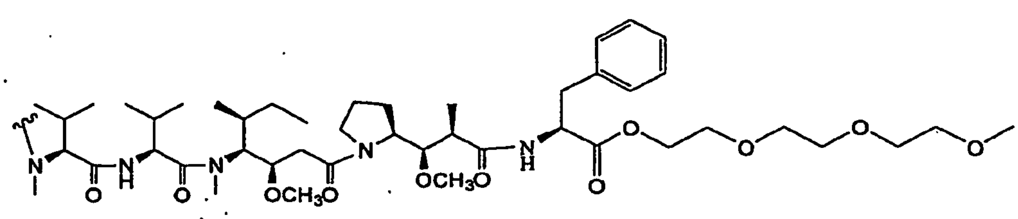
[00195] Illustrative Drug Units D_E and D_F include the units having the following structures:

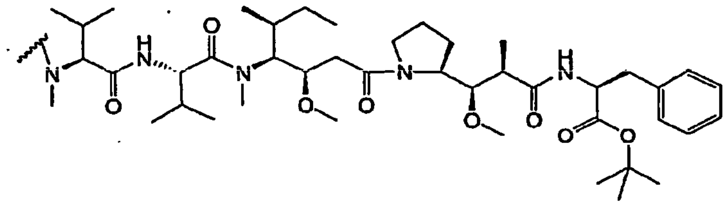


MMAE

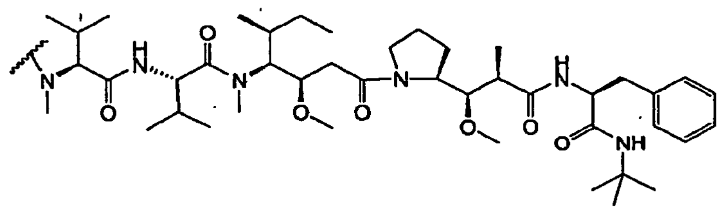


MMAF

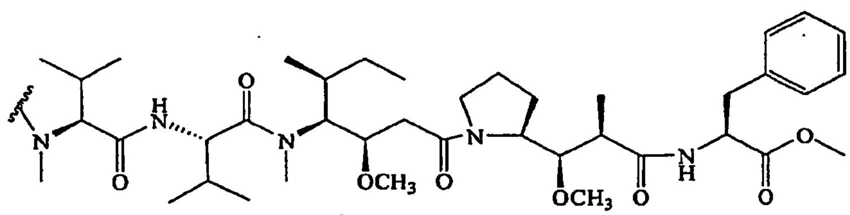



N
H
N
O
O
N
O
N
O
O
O
H
N
O
O
O

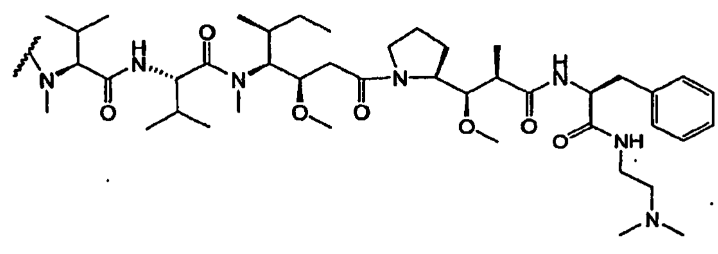
,


N
H
N
O
O
N
O
N
O
O
O
H
N
O
NH

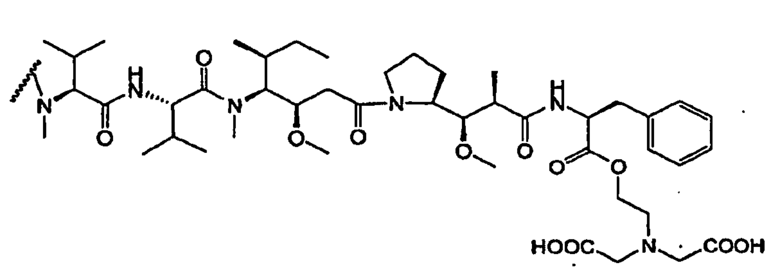
,


N
H
N
O
O
N
OCH3
O
N
OCH3
O
N
H
O
O

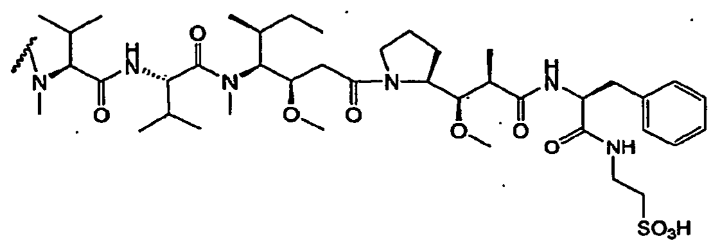
,


N
H
N
O
O
N
O
N
O
O
O
H
N
O
NH
N

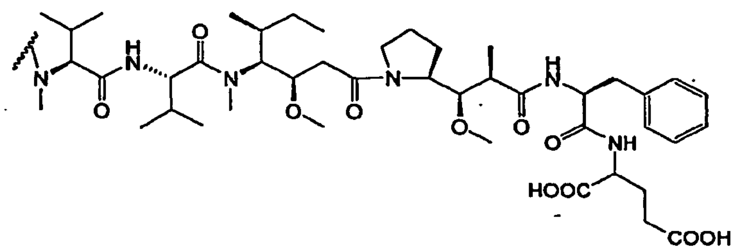
,


N
H
N
O
O
N
O
N
O
O
O
H
N
O
O
HOOC
N
COOH

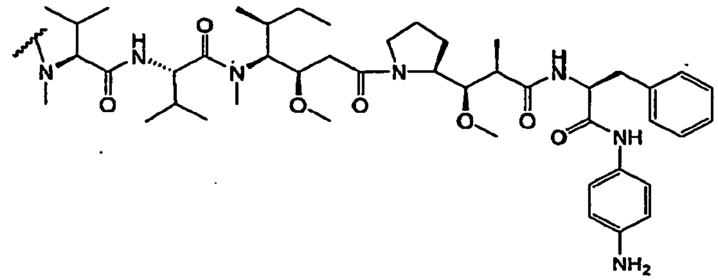
,



,



, and

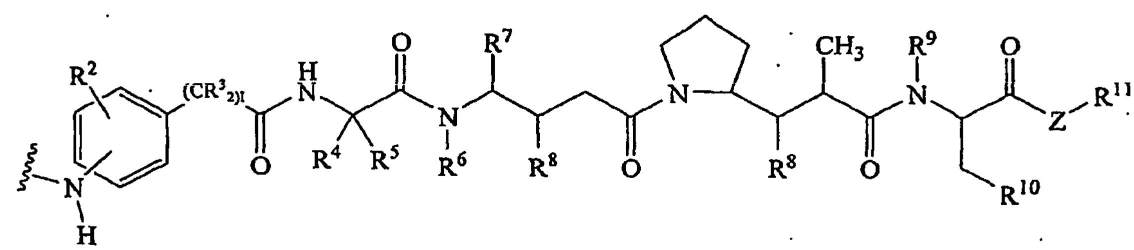


:

and pharmaceutically acceptable salts or solvates thereof.

[00196] In one aspect, hydrophilic groups, such as but not limited to triethylene glycol esters (TEG), as shown above, can be attached to the Drug Unit at R11. Without being bound by theory, the hydrophilic groups assist in the internalization and non-agglomeration of the Drug Unit.

[00197] In another aspect, the Drug Unit is an amino-benzoic acid derivative of an auristatin of the following formula:



wherein, independently at each location:

R^2 is selected from –hydrogen –C_1-C_8 alkyl, -O-(C_1-C_8 alkyl), -halogen, -NO_2, -COOH, and -C(O)OR^{11};

each R^3 is selected independently from –hydrogen and –C_1-C_8 alkyl;

l is an integer ranging from 0-10;

R^4 is selected from -hydrogen, -C_1-C_8 alkyl, -C_3-C_8 carbocycle, -aryl, -C_1-C_8 alkyl-aryl, -C_1-C_{10} alkyl-(C_3-C_8 carbocycle), -C_3-C_8 heterocycle and -C_1-C_{10} alkyl-(C_3-C_8 heterocycle), and R^5 is selected from -H and –methyl; or R^4 and R^5 jointly have the formula -$(CR^aR^b)_n$-, wherein R^a and R^b are independently selected from -H, -C_1-C_8 alkyl and -C_3-C_8 carbocycle and n is selected from 2, 3, 4, 5 and 6, and form a ring with the carbon atom to which they are attached;

R^6 is selected from -H and -C_1-C_8 alkyl;

R^7 is selected from -H, -C_1-C_8 alkyl, -C_3-C_8 carbocycle, aryl, -C_1-C_{10} alkyl-aryl, -C_1-C_{10} alkyl-(C_3-C_8 carbocycle), -C_3-C_8 heterocycle and -C_1-C_{10} alkyl-(C_3-C_8 heterocycle);

each R^8 is independently selected from -H, -OH, -C_1-C_8 alkyl, -C_3-C_8 carbocycle, -O-alkyl-(C_1-C_8 carbocycle) and -O-(C_1-C_8 alkyl);

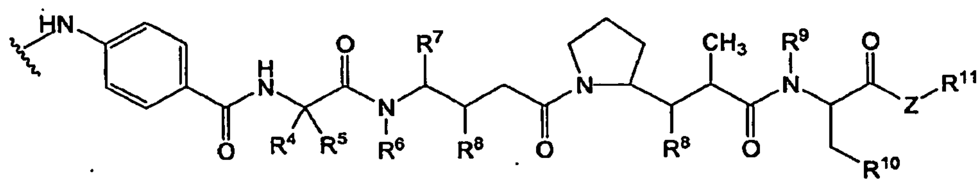
R^9 is selected from -H and -C_1-C_8 alkyl;

R^{10} is selected from aryl or -C_3-C_8 heterocycle;

Z is -O-, -S-, -NH-, or -NR^{12}- where R^{12} is C_1-C_8 alkyl or aryl; and

R^{11} is selected from –H, C_1-C_8 alkyl, aryl, -C_3-C_8 heterocycle, -$(CH_2CH_2O)_r$-H, -$(CH_2CH_2O)_r$-CH_3, and -$(CH_2CH_2O)_r$-$CH_2CH_2C(O)OH$; wherein r is an integer ranging from 1-10.

[00198] In some embodiments, the Drug Unit is of the following formula:



wherein, independently at each location:

R^4 is selected from -hydrogen, -C_1-C_8 alkyl, -C_3-C_8 carbocycle, -aryl, -C_1-C_{10} alkyl-aryl, -C_1-C_{10} alkyl-(C_3-C_8 carbocycle), -C_3-C_8 heterocycle and -C_1-C_{10} alkyl-(C_3-C_8 heterocycle), and R^5 is selected from -H and –methyl; or R^4 and R^5 jointly have the formula -$(CR^aR^b)_n$-, wherein R^a and R^b are independently selected from -H, -C_1-C_8 alkyl and -C_3-C_8 carbocycle and n is selected from 2, 3, 4, 5 and 6, and form a ring with the carbon atom to which they are attached;

R^6 is selected from -H and -C_1-C_8 alkyl;

R^7 is selected from -H, -C_1-C_8 alkyl, -C_3-C_8 carbocycle, aryl, -C_1-C_{10} alkyl-aryl, -C_1-C_{10} alkyl-(C_3-C_8 carbocycle), -C_3-C_8 heterocycle and -C_1-C_{10} alkyl-(C_3-C_8 heterocycle);

each R^8 is independently selected from -H, -OH, -C_1-C_8 alkyl, -C_3-C_8 carbocycle, -O-alkyl-(C_1-C_8 carbocycle) and -O-(C_1-C_8 alkyl);

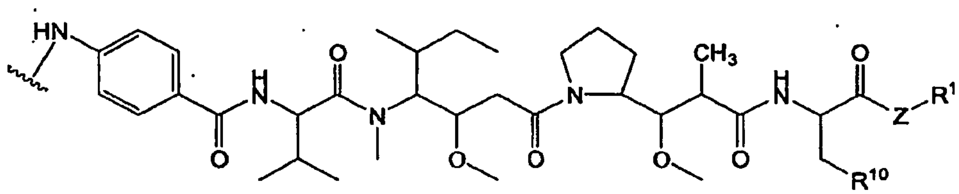
R^9 is selected from -H and -C_1-C_8 alkyl;

R^{10} is selected from aryl or -C_3-C_8.heterocycle;

Z is -O-, -S-, -NH-, or -NR^{12}- where R^{12} is C_1-C_8 alkyl or aryl; and

R^{11} is selected from –H, C_1-C_8 alkyl, aryl, -C_3-C_8 heterocycle, -$(CH_2CH_2O)_r$-H, -$(CH_2CH_2O)_r$-CH_3, and -$(CH_2CH_2O)_r$-$CH_2CH_2C(O)OH$; wherein r is an integer ranging from 1-10.

[00199] In some embodiments, the Drug Unit is of the following formula:



wherein, independently at each location:

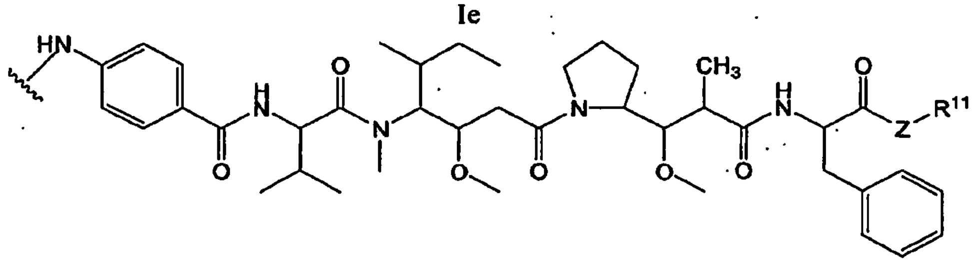
R^{10} is selected from aryl group or -C_3-C_8 heterocycle;

Z is -O-, -S-, -NH-, or -NR^{12}- where R^{12} is C_1-C_8 alkyl or aryl; and

R^{11} is selected from –H, C_1-C_8 alkyl, aryl, -C_3-C_8 heterocycle, -$(CH_2CH_2O)_r$-H, -$(CH_2CH_2O)_r$-CH_3, and -$(CH_2CH_2O)_r$-$CH_2CH_2C(O)OH$; wherein r is an integer ranging from 1-10.

[00200] In some embodiments, the Drug Unit is of the following formula:

Ie

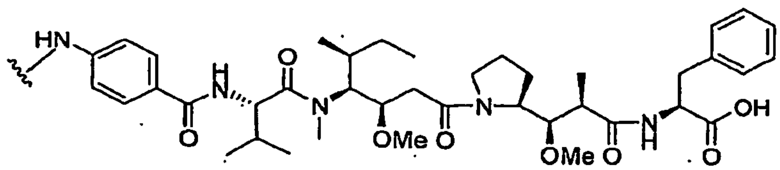


wherein:

Z is -O-, -S-, -NH-, or -NR^{12}- where R^{12} is C_1-C_8 alkyl or aryl; and

R^{11} is selected from –H, C_1-C_8 alkyl, aryl, -C_3-C_8 heterocycle, -$(CH_2CH_2O)_r$-H, -$(CH_2CH_2O)_r$-CH_3, and -$(CH_2CH_2O)_r$-$CH_2CH_2C(O)OH$; wherein r is an integer ranging from 1-10.

[00201] In some embodiments, the Drug Unit is of the following formula:



4. Prominin-1 Nucleic Acid Molecules

[00202] Isolated prominin-1 nucleic acid molecules of the present invention consist of, consist essentially of, or comprise a nucleotide sequence that encodes a prominin-1 protein of the present invention, an allelic variant thereof, or an ortholog or paralog thereof. As used herein, an "isolated" nucleic acid molecule is one that is separated from other nucleic acid present in the natural source of the nucleic acid. Preferably, an "isolated" nucleic acid is free of sequences which naturally flank the nucleic acid (i.e., sequences located at the 5' and 3' ends of the nucleic acid) in the genomic DNA of the organism from which the nucleic acid is derived. However, there can be some flanking nucleotide sequences, for example up to about 5 kilobases (KB), 4KB, 3KB, 2KB, or 1KB or less; particularly contiguous peptide encoding sequences and peptide encoding sequences within the same gene but separated by introns in the genomic sequence. The important point is that the nucleic acid is isolated from remote and unimportant flanking sequences such that it can be subjected to the specific manipulations described herein such as recombinant expression, preparation of probes and primers, and other uses specific to the nucleic acid sequences.

[00203] Moreover, an "isolated" nucleic acid molecule, such as a transcript/cDNA molecule,. can be substantially free of other cellular material, or culture medium when produced by recombinant techniques, or chemical precursors or other chemicals when chemically synthesized. However, the nucleic acid molecule can be fused to other coding or regulatory sequences and still be considered isolated.

[00204] For example, recombinant DNA molecules contained in a vector are considered isolated. Further examples of isolated DNA molecules include recombinant DNA

molecules maintained in heterologous host cells or purified (partially or substantially) DNA molecules in solution. Isolated RNA molecules include in vivo or in vitro RNA transcripts of the isolated DNA molecules of the present invention. Isolated nucleic acid molecules according to the present invention further include such molecules produced synthetically.

[00205] The isolated nucleic acid molecules can encode the mature protein plus additional amino or carboxyl-terminal amino acids, or amino acids interior to the mature peptide (when the mature form has more than one peptide chain, for instance). Such sequences may play a role in processing of a protein from precursor to a mature form, facilitate protein trafficking, prolong or shorten protein half-life or facilitate manipulation of a protein for assay or production, among other things. As generally is the case in situ, the additional amino acids may be processed away from the mature protein by cellular enzymes.

[00206] As mentioned above, the isolated nucleic acid molecules include, but are not limited to, the sequence encoding prominin-1 peptide alone, the sequence encoding the mature peptide and additional coding sequences, such as a leader or secretory sequence (e.g., a pre-pro or pro-protein sequence), the sequence encoding the mature peptide, with or without the additional coding sequences, plus additional non-coding sequences, for example introns and non-coding 5' and 3' sequences such as transcribed but non-translated sequences that play a role in transcription, mRNA processing (including splicing and polyadenylation signals), ribosome binding and stability of mRNA. In addition, the nucleic acid molecule may be fused to a marker sequence encoding, for example, a peptide that facilitates purification.

[00207] Isolated nucleic acid molecules can be in the form of RNA, such as mRNA, or in the form DNA, including cDNA and genomic DNA obtained by cloning or produced by chemical synthetic techniques or by a combination thereof. The nucleic acid, especially DNA, can be double-stranded or single-stranded. Single-stranded nucleic acid can be the coding strand (sense strand) or the non-coding strand (anti-sense strand).

[00208] The invention further provides nucleic acid molecules that encode fragments of the proteins of the present invention as well as nucleic acid molecules that encode obvious

variants of prominin-1 protein of the present invention that are described above. Such nucleic acid molecules may be naturally occurring, such as allelic variants (same locus), paralogs (different locus), and orthologs (different organism), or may be constructed by recombinant DNA methods or by chemical synthesis. Such non-naturally occurring variants may be made by mutagenesis techniques, including those applied to nucleic acid molecules, cells, or organisms. Accordingly, as discussed above, the variants can contain nucleotide substitutions, deletions, inversions and insertions. Variation can occur in either or both the coding and non-coding regions. The variations can produce both conservative and non-conservative amino acid substitutions.

[00209] A fragment comprises a contiguous nucleotide sequence greater than 12 or more nucleotides. Further, a fragment could be at least 30, 40, 50, 100, 250 or 500 nucleotides in length. The length of the fragment will be based on its intended use. For example, the fragment can encode epitope bearing regions of the peptide, or can be useful as DNA probes and primers. Such fragments can be isolated using the known nucleotide sequence to synthesize an oligonucleotide probe. A labeled probe can then be used to screen a cDNA library, genomic DNA library, or mRNA to isolate nucleic acid corresponding to the coding region. Further, primers can be used in PCR reactions to clone specific regions of gene.

[00210] A probe/primer typically comprises substantially a purified oligonucleotide or oligonucleotide pair. The oligonucleotide typically comprises a region of nucleotide sequence that hybridizes under stringent conditions to at least about 12, 20, 25, 40, 50 or more consecutive nucleotides.

[00211] Orthologs, homologs, and allelic variants can be identified using methods well known in the art. As described in the Section 1 (supra), these variants comprise a nucleotide sequence encoding a peptide that is typically 60-70%, 70-80%, 80-90%, and more typically at least about 90-95% or more homologous to the nucleotide sequence. Such nucleic acid molecules can readily be identified as being able to hybridize under moderate to stringent conditions, to the nucleotide sequence shown in the Sequence Listing

or a fragment of the sequence. Allelic variants can readily be determined by genetic locus of the encoding gene.

[00212] As used herein, the term "hybridizes under stringent conditions" is intended to describe conditions for hybridization and washing under which nucleotide sequences encoding a peptide at least 60-70% homologous to each other typically remain hybridized to each other. The conditions can be such that sequences at least about 60%, at least about 70%, or at least about 80% or more homologous to each other typically remain hybridized to each other. Such stringent conditions are known to those skilled in the art and can be found in Current Protocols in Molecular Biology, John Wiley & Sons, N.Y. (1989-2006), 6.3.1-6.3.6. One example of stringent hybridization conditions is hybridization in 6X sodium chloride/sodium citrate (SSC) at about 45°C, followed by one or more washes in 0.2 X SSC, 0.1% SDS at 50-65C. Examples of moderate to low stringency hybridization conditions are well known in the art.

5. Vectors and Host Cells

[00213] The invention also provides vectors containing the nucleic acid molecules described herein. The term "vector" refers to a vehicle, preferably a nucleic acid molecule, which can transport the nucleic acid molecules. When the vector is a nucleic acid molecule, the nucleic acid molecules are covalently linked to the vector nucleic acid. With this aspect of the invention, the vector includes a plasmid, single or double stranded phage, a single or double stranded RNA or DNA viral vector, or artificial chromosome, such as a BAC, PAC, YAC, OR MAC.

[00214] A vector can be maintained in the host cell as an extrachromosomal element where it replicates and produces additional copies of the nucleic acid molecules. Alternatively, the vector may integrate into the host cell genome and produce additional copies of the nucleic acid molecules when the host cell replicates.

[00215] The invention provides vectors for the maintenance (cloning vectors) or vectors for expression (expression vectors) of the nucleic acid molecules. The vectors can function in prokaryotic or eukaryotic cells or in both (shuttle vectors).

[00216] Expression vectors contain cis-acting regulatory regions that are operably linked in the vector to the nucleic acid molecules such that transcription of the nucleic acid molecules is allowed in a host cell. The nucleic acid molecules can be introduced into the host cell with a separate nucleic acid molecule capable of affecting transcription. Thus, the second nucleic acid molecule may provide a trans-acting factor interacting with the cis-regulatory control region to allow transcription of the nucleic acid molecules from the vector. Alternatively, a trans-acting factor may be supplied by the host cell. Finally, a trans-acting factor can be produced from the vector itself. It is understood, however, that in some embodiments, transcription and/or translation of the nucleic acid molecules can occur in a cell-free system.

[00217] The regulatory sequences to which the nucleic acid molecules described herein can be operably linked include promoters for directing mRNA transcription. These include, but are not limited to, the left promoter from bacteriophage, the lac, TRP, and TAC promoters from *E. coli*, the early and late promoters from SV40, the CMV immediate early promoter, the adenovirus early and late promoters, and retrovirus long-terminal repeats.

[00218] In addition to control regions that promote transcription, expression vectors may also include regions that modulate transcription, such as repressor binding sites and enhancers. Examples include the SV40 enhancer, the cytomegalovirus immediate early enhancer, polyoma enhancer, adenovirus enhancers, and retrovirus LTR enhancers.

[00219] In addition to containing sites for transcription initiation and control, expression vectors can also contain sequences necessary for transcription termination and, in the transcribed region a ribosome binding site for translation. Other regulatory control elements for expression include initiation and termination codons as well as polyadenylation signals. The person of ordinary skill in the art would be aware of the numerous regulatory sequences that are useful in expression vectors. Such regulatory sequences are described, for example, in Sambrook et al., Molecular Cloning: A Laboratory Manual. 3rd. ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY (2001).

[00220] A variety of expression vectors can be used to express a nucleic acid molecule. Such vectors include chromosomal, episomal, and virus-derived vectors, for example vectors derived from bacterial plasmids, from bacteriophage, from yeast episomes, from yeast chromosomal elements, including yeast artificial chromosomes, from viruses such as baculoviruses, papovaviruses such as SV40, Vaccinia viruses, adenoviruses, poxviruses, pseudorabies viruses, and retroviruses. Vectors may also be derived from combinations of these sources such as those derived from plasmid and bacteriophage genetic elements, e.g. cosmids and phagemids. Appropriate cloning and expression vectors for prokaryotic and eukaryotic hosts are described in Sambrook et al., Molecular Cloning: A Laboratory Manual. 3rd. ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY (2001).

[00221] The regulatory sequence may provide constitutive expression in one or more host cells (i.e. tissue specific) or may provide for inducible expression in one or more cell types such as by temperature, nutrient additive, or exogenous factor such as a hormone or other ligand. A variety of vectors providing for constitutive and inducible expression in prokaryotic and eukaryotic hosts are well known to those of ordinary skill in the art.

[00222] The nucleic acid molecules can be inserted into the vector nucleic acid by well-known methodology. Generally, the DNA sequence that will ultimately be expressed is joined to an expression vector by cleaving the DNA sequence and the expression vector with one or more restriction enzymes and then ligating the fragments together. Procedures for restriction enzyme digestion and ligation are well known to those of ordinary skill in the art.

[00223] The vector containing the appropriate nucleic acid molecule can be introduced into an appropriate host cell for propagation or expression using well-known techniques. Bacterial cells include, but are not limited to, *E. coli*, *Streptomyces*, and *Salmonella typhimurium*. Eukaryotic cells include, but are not limited to, yeast, insect cells such as *Drosophila*, animal cells such as COS and CHO cells (e.g., DG44 or CHO-s), and plant cells.

[00224] As described herein, it may be desirable to express the peptide as a fusion protein. Accordingly, the invention provides fusion vectors that allow for the production of the

peptides. Fusion vectors can increase the expression of a recombinant protein; increase the solubility of the recombinant protein, and aid in the purification of the protein by acting for example as a ligand for affinity purification. A proteolytic cleavage site may be introduced at the junction of the fusion moiety so that the desired peptide can ultimately be separated from the fusion moiety. Proteolytic enzymes include, but are not limited to, factor Xa, thrombin, and enteroenzyme. Typical fusion expression vectors include pGEX (Smith et al., Gene 67:31-40 (1988)), pMAL (New England Biolabs, Beverly, MA) and pRIT5 (Pharmacia, Piscataway, NJ) which fuse glutathione S-transferase (GST), maltose E binding protein, or protein A, respectively, to the target recombinant protein. Examples of suitable inducible non-fusion *E. coli* expression vectors include pTrc (Amann et al., Gene 69:301-315 (1988)) and pET 11d (Studier et al., Gene Expression Technology: Methods in Enzymology 185:60-89 (1990)).

[00225] Recombinant protein expression can be maximized in host bacteria by providing a genetic background wherein the host cell has an impaired capacity to proteolytically cleave the recombinant protein. (Gottesman, S., Gene Expression Technology: Methods in Enzymology 185, Academic Press, San Diego, California (1990), pp. 119-128). Alternatively, the sequence of the nucleic acid molecule of interest can be altered to provide preferential codon usage for a specific host cell, for example *E. coli*. (Wada et al., Nucleic Acids Res. 20:2111-2118 (1992)).

[00226] The nucleic acid molecules can also be expressed by expression vectors suitable in a yeast host. Examples of vectors for expression in yeast e.g., *S. cerevisiae* include pYepSec1 (Baldari, et al., EMBO J. 6:229-234 (1987)), pMFa (Kurjan et al., Cell 30:933-943 (1982)), pJRY88 (Schultz et al., Gene 54:113-123 (1987)), and pYES2 (Invitrogen Corporation, San Diego, CA).

[00227] The nucleic acid molecules can also be expressed in insect cells using, for example, baculovirus expression vectors. Baculovirus vectors available for expression of proteins in cultured insect cells (e.g., Sf 9 cells) include the pAc series (Smith et al., Mol. Cell Biol. 3:2156-2165 (1983)) and the pVL series (Lucklow et al., Virology 170:31-39 (1989)).

[00228] In certain embodiments of the invention, the nucleic acid molecules described herein are expressed in mammalian cells using mammalian expression vectors. Examples of mammalian expression vectors include pCDM8 (Seed, B. Nature 329:840 (1987)), pMT2PC (Kaufman et al., EMBO J. 6:187-195 (1987)), andCHEF (U.S. Pat. No. 5,888,809).

[00229] The expression vectors listed herein are provided by way of example only of the well-known vectors available to those of ordinary skill in the art that would be useful to express the nucleic acid molecules. The person of ordinary skill in the art would be aware of other vectors suitable for maintenance propagation or expression of the nucleic acid molecules described herein. These are found for example in Sambrook, J., Fritsh, E. F., and Maniatis, T. Molecular Cloning: A Laboratory Manual. 3rd. ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY (2001).

[00230] The invention also encompasses vectors in which the nucleic acid sequences described herein are cloned into the vector in reverse orientation, but operably linked to a regulatory sequence that permits transcription of antisense RNA. Thus, an antisense transcript can be produced to all, or to a portion, of the nucleic acid molecule sequences described herein, including both coding and non-coding regions. Expression of this antisense RNA is subject to each of the parameters described above in relation to expression of the sense RNA (e.g., regulatory sequences, constitutive or inducible expression, tissue-specific expression).

[00231] The invention also relates to recombinant host cells containing the vectors described herein. Host cells therefore include prokaryotic cells, lower eukaryotic cells such as yeast, other eukaryotic cells such as insect cells, and higher eukaryotic cells such as mammalian cells.

[00232] The recombinant host cells are prepared by introducing the vector constructs described herein into the cells by techniques readily available to the person of ordinary skill in the art. These include, but are not limited to, calcium phosphate transfection, DEAE-dextran-mediated transfection, cationic lipid-mediated transfection, electroporation, transduction, infection, lipofection, and other techniques such as those found in Sambrook,

et al. (Molecular Cloning: A Laboratory Manual. 3rd. ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY (2001).

[00233] Host cells can contain more than one vector. Thus, different nucleotide sequences can be introduced on different vectors of the same cell. Similarly, the nucleic acid molecules can be introduced either alone or with other nucleic acid molecules that are not related to the nucleic acid molecules such as those providing trans-acting factors for expression vectors. When more than one vector is introduced into a cell, the vectors can be introduced independently, co-introduced or joined to the nucleic acid molecule vector.

[00234] In the case of bacteriophage and viral vectors, these can be introduced into cells as packaged or encapsulated virus by standard procedures for infection and transduction. Viral vectors can be replication-competent or replication-defective. In the case in which viral replication is defective, replication will occur in host cells providing functions that complement the defects.

[00235] Vectors generally include selectable markers that enable the selection of the subpopulation of cells that contain the recombinant vector constructs. The marker can be contained in the same vector that contains the nucleic acid molecules described herein or may be on a separate vector. Markers include tetracycline or ampicillin-resistance genes for prokaryotic host cells and dihydrofolate reductase or neomycin resistance for eukaryotic host cells. However, any marker that provides selection for a phenotypic trait will be effective.

[00236] While the mature proteins can be produced in bacteria, yeast, mammalian cells, and other cells under the control of the appropriate regulatory sequences, cell- free transcription and translation systems can also be used to produce these proteins using RNA derived from the DNA constructs described herein.

[00237] Where secretion of the peptide is desired, which may be difficult to achieve with a multi-transmembrane domain containing protein such as prominin-1, appropriate secretion signals are incorporated into the vector. The signal sequence can be endogenous to the peptides or heterologous to these peptides.

[00238] Where the peptide is not secreted into the medium, the protein can be isolated from the host cell by standard disruption procedures, including freeze thaw, sonication, mechanical disruption, use of lysing agents and the like. The peptide can then be recovered and purified by well-known purification methods including ammonium sulfate precipitation, acid extraction, anion or cationic exchange chromatography, phosphocellulose chromatography, hydrophobic-interaction chromatography, affinity chromatography, hydroxylapatite chromatography, lectin chromatography, or high performance liquid chromatography.

[00239] It is also understood that depending upon the host cell in recombinant production of the peptides described herein, the peptides can have various glycosylation patterns, depending upon the cell, or maybe non-glycosylated as when produced in bacteria. In addition, the peptides may include an initial modified methionine in some cases as a result of a host-mediated process.

[00240] The recombinant host cells expressing the peptides described herein have a variety of uses. First, the cells are useful for producing prominin-1 protein or peptide that can be further purified to produce desired amounts of prominin-1 protein or fragments. Thus, host cells containing expression vectors are useful for peptide production.

[00241] Host cells are also useful for conducting cell-based assays involving the prominin-1 protein or prominin-1 protein fragments, such as those described above as well as other formats known in the art. Thus, a recombinant host cell expressing a native prominin-1 protein is useful for assaying compounds that stimulate or inhibit prominin-1 protein function.

[00242] Host cells are also useful for identifying prominin-1 protein mutants in which these functions are affected. If the mutants naturally occur and give rise to a pathology, host cells containing the mutations are useful to assay compounds that have a desired effect on the mutant prominin-1 protein (for example, stimulating or inhibiting function) which may not be indicated by their effect on the native prominin-1 protein.

6. Detection and Diagnosis in General

[00243] As used herein, a “biological sample” can be collected from tissues, blood, sera, cell lines or biological fluids such as, plasma, interstitial fluid, urine, cerebrospinal fluid, and the like, containing cells. In preferred embodiments, a biological sample comprises cells or tissues suspected of having diseases (e.g., cells obtained from a biopsy).

[00244] As used herein, a “differential level” is defined as the level of prominin-1 protein or nucleic acids in a test sample either above or below the level in control samples, wherein the level of control samples is obtained either from a control cell line, a normal tissue or body fluid(s), or combination thereof, from a healthy subject. While the protein is overexpressed, the expression of prominin-1 is preferably greater than about 20%, or preferably greater than about 30%, and most preferably greater than about 50% or more of disease sample, at a level that is at least two fold, and preferably at least five fold, greater than the level of expression in control samples, as determined using a representative assay provided herein. While the protein is under expressed, the expression of prominin-1 is preferably less than about 20%, or preferably less than 30%, and most preferably less than about 50% or more of the disease sample, at a level that is at least 0.5 fold, and preferably at least 0.2 fold less than the level of the expression in control samples, as determined using a representative assay provided herein.

[00245] As used herein, a “subject” can be a mammalian subject or non mammalian subject, preferably, a mammalian subject. A mammalian subject can be human or non-human, preferably human. A healthy subject is defined as a subject without detectable diseases or associated pathologies by using conventional diagnostic methods.

[00246] As used herein, the “disease(s)” preferably include cancer, particularly breast, bladder, colon, colorectal, kidney, liver, lung, melanoma, ovary, pancreatic, pharyngeal, gastrointestinal (e.g., gastric or colorectal), glioblastoma, and prostate cancer and associated diseases and pathologies.

7. Treatment in General

[00247] This invention further pertains to novel agents identified by the screening assays described below. It is also within the scope of this invention to use an agent identified for treatment purposes. For example, an agent identified as described herein (e.g., a prominin-1-modulating agent, an antisense prominin-1 nucleic acid molecule, a prominin-1-RNAi fragment, a prominin-1-specific antibody, a prominin-1-specific antibody-drug conjugate, or a prominin-1-binding partner) can be used in an animal or other model to determine the efficacy, toxicity, or side effects of treatment with such an agent. Alternatively, an agent identified as described herein can be used in an animal or other model to determine the mechanism of action of such an agent. Furthermore, this invention pertains to uses of novel agents identified by the above-described screening assays for treatments as described herein.

[00248] Modulators of prominin-1 protein activity identified according to these drug screening assays can be used to treat a subject with a disorder mediated by prominin-1, e.g., by treating cells or tissues that express prominin-1 at a differential level. Methods of treatment include the steps of administering a modulator of prominin-1 activity in a pharmaceutical composition to a subject in need of such treatment.

[00249] The following terms, as used in the present specification and claims, are intended to have the meaning as defined below, unless indicated otherwise.

[00250] "Treat," "treating" or "treatment" of a disease includes: (1) inhibiting the disease, i.e., arresting or reducing the development of the disease or its clinical symptoms, or (2) relieving the disease, i.e., causing regression of the disease or its clinical symptom(s).

[00251] The term "prophylaxis" is used to distinguish from "treatment," and to encompass both "preventing" and "suppressing." It is not always possible to distinguish between "preventing" and "suppressing," as the ultimate inductive event or events may be unknown, latent, or the patient is not ascertained until well after the occurrence of the event or events. Therefore, the term "protection," as used herein, is meant to include "prophylaxis."

[00252] A "therapeutically effective amount" means the amount of an agent that, when administered to a subject for treating a disease, is sufficient to effect such treatment for the disease. The "therapeutically effective amount" will vary depending on the agent, the disease and its severity and the age, weight, etc., of the subject to be treated.

[00253] In one embodiment, when decreased expression or activity of the protein is desired, an inhibitor, antagonist, antibody and the like or a pharmaceutical agent containing one or more of these molecules may be delivered. Such delivery may be effected by methods well known in the art and may include delivery by an antibody specifically targeted to the protein.

[00254] In another embodiment, when increased expression or activity of the protein is desired, the protein, an agonist, an enhancer and the like or a pharmaceutical agent containing one or more of these molecules may be delivered. Such delivery may be effected by methods well known in the art.

[00255] While it is possible for the modulating agent to be administered in a pure or substantially pure form, it is preferable to present it as a pharmaceutical composition, formulation or preparation with a carrier. The formulations of the present invention, both for veterinary and for human use, comprise a suitable active prominin-1 modulating agent, together with one or more pharmaceutically acceptable carriers and, optionally, other therapeutic ingredients. The carrier(s) must be "acceptable" in the sense of being compatible with the other ingredients of the formulation and not deleterious to the recipient thereof. The formulations may conveniently be presented in unit dosage form and may be prepared by any method well-known in the pharmaceutical art.

[00256] Suitable pharmaceutical carriers include proteins such as albumins (e.g., U.S. Pat. No. 4,507,234, to Kato et al.), peptides and polysaccharides such as aminodextran (e.g., U.S. Pat. No. 4,699,784, to Shih et al.), or water. A carrier may also bear an agent by noncovalent bonding or by encapsulation, such as within a liposome vesicle (e.g., U.S. Pat. Nos. 4,429,008 and 4,873,088). Carriers specific for radionuclide agents include radiohalogenated small molecules and chelating compounds. For example, U.S. Pat. No. 4,735,792 discloses representative radiohalogenated small molecules and their synthesis. A

radionuclide chelate may be formed from chelating compounds that include those containing nitrogen and sulfur atoms as the donor atoms for binding the metal, metal oxide, radionuclide. For example, U.S. Pat. No. 4,673,562, to Davison et al. discloses representative chelating compounds and their synthesis.

[00257] All methods include the step of bringing into association the active ingredient with the carrier, which constitutes one or more accessory ingredients. In general, the formulations are prepared by uniformly and intimately bringing into association the active ingredient with liquid carriers or finely divided solid carriers or both, and then, if necessary, shaping the product into the desired formulation.

[00258] Formulations suitable for intravenous, intramuscular, subcutaneous, or intraperitoneal administration conveniently comprise sterile aqueous solutions of the active ingredient with solutions, which are preferably isotonic with the blood of the recipient. Such formulations may be conveniently prepared by dissolving solid active ingredient in water containing physiologically compatible substances such as sodium chloride (e.g., 0.1-2.0M), glycine, and the like, and having a buffered pH compatible with physiological conditions to produce an aqueous solution, and rendering said solution sterile. These may be present in unit or multi-dose containers, for example, sealed ampoules or vials.

[00259] The formulations of the present invention may incorporate a stabilizer. Illustrative stabilizers are polyethylene glycol, proteins, saccharides, amino acids, inorganic acids, detergents, and organic acids, which may be used either on their own or as admixtures. These stabilizers are preferably incorporated in an amount of 0.11-10,000 parts by weight per part by weight of an agent. If two or more stabilizers are to be used, their total amount is preferably within the range specified above. These stabilizers are used in aqueous solutions at the appropriate concentration and pH. The specific osmotic pressure of such aqueous solutions is generally in the range of 0.1-3.0 osmoles, preferably in the range of 0.8-1.2. The pH of the aqueous solution is adjusted to be within the range of 5.0-9.0, preferably within the range of 6-8. In formulating an antibody or antibody drug conjugate, an anti-adsorption agent may be used.

[00260] Additional pharmaceutical methods may be employed to control the duration of action. Controlled release preparations may be achieved through the use of polymer to complex or absorb the proteins or their derivatives. The controlled delivery may be exercised by selecting appropriate macromolecules (for example polyester, polyamino acids, polyvinyl, pyrrolidone, ethylenevinylacetate, methylcellulose, carboxymethylcellulose, or protamine sulfate) and the concentration of macromolecules as well as the methods of incorporation in order to control release. Another possible method to control the duration of action by controlled-release preparations is to incorporate anti-prominin-1 antibody into particles of a polymeric material such as polyesters, polyamino acids, hydrogels, poly(lactic acid) or ethylene vinylacetate copolymers. Alternatively, instead of incorporating these agents into polymeric particles, it is possible to entrap these materials in microcapsules prepared, for example, by coacervation techniques or by interfacial polymerization, for example, hydroxymethylcellulose or gelatin-microcapsules and poly(methylmethacylate) microcapsules, respectively, or in colloidal drug delivery systems, for example, liposomes, albumin microspheres, microemulsions, nanoparticles, and nanocapsules or in macroemulsions.

[00261] When oral preparations are desired, the compositions may be combined with typical carriers, such as lactose, sucrose, starch, talc magnesium stearate, crystalline cellulose, methyl cellulose, carboxymethyl cellulose, glycerin, sodium alginate or gum arabic among others.

8. Diagnosis, Treatment and Screening Methods Using Prominin-1 Nucleic Acids

a. General Aspects

[00262] The nucleic acid molecules of the present invention are useful for probes, primers, chemical intermediates, and in biological assays. The nucleic acid molecules are useful as a hybridization probe for messenger RNA, transcript/cDNA and genomic DNA to detect or isolate full-length cDNA and genomic clones encoding prominin-1 protein or peptide of the invention, or variants thereof

[00263] The probe can correspond to any sequence along the entire length of a nucleic acid molecule of SEQ ID NOS:6-10. Accordingly, it could be derived from 5' noncoding regions, the coding region, and 3' noncoding regions.

[00264] The nucleic acid molecules are also useful as primers for PCR to amplify any given region of a nucleic acid molecule and are useful to synthesize antisense molecules of desired length and sequence.

[00265] The nucleic acid molecules are also useful for constructing recombinant vectors. Such vectors include expression vectors that express a portion of, or all of, the peptide sequences. The nucleic acid molecules are also useful for expressing antigenic portions of the proteins.

[00266] The nucleic acid molecules are also useful for designing ribozymes corresponding to all, or a part, of the mRNA produced from the nucleic acid molecules described herein.

[00267] The nucleic acid molecules are also useful for constructing host cells expressing a part, or all, of the nucleic acid molecules and peptides.

[00268] The nucleic acid molecules are also useful for constructing transgenic animals expressing all, or a part, of the nucleic acid molecules and peptides.

[00269] In vitro techniques for detection of mRNA include Northern hybridizations and in situ hybridizations. In vitro techniques for detecting DNA include Southern hybridizations and in situ hybridization.

b. Diagnosis Methods

[00270] The nucleic acid molecules are also useful as hybridization probes for determining the presence, level, form and distribution of nucleic acid expression. The probes can be used to detect the presence of, or to determine levels of, a specific nucleic acid molecule in cells, tissues, and in organisms. Accordingly, probes corresponding to the peptides described herein can be used to assess expression and/or gene copy number in a given cell, tissue, or organism. These uses are relevant for diagnosis of disorders involving an increase or decrease in prominin-1 protein expression relative to normal results.

[00271] Probes can be used as a part of a diagnostic test kit for identifying cells or tissues that express prominin-1 protein differentially, such as by measuring a level of a prominin-1-encoding nucleic acid in a sample of cells from a subject e.g., mRNA or genomic DNA, or determining if a prominin-1 gene has been mutated.

[00272] The invention also encompasses kits for detecting the presence of prominin-1 nucleic acid in a biological sample. For example, the kit can comprise reagents such as a labeled or labelable nucleic acid or agent capable of detecting prominin-1 nucleic acid in a biological sample; means for determining the amount of prominin-1 nucleic acid in the sample; and means for comparing the amount of prominin-1 nucleic acid in the sample with a standard. The compound or agent can be packaged in a suitable container. The kit can further comprise instructions for using the kit to detect prominin-1 protein mRNA or DNA.

c. Screening Method Using Nucleic Acids

[00273] Nucleic acid expression assays are useful for drug screening to identify compounds that modulate prominin-1 nucleic acid expression.

[00274] The invention thus provides a method for identifying a compound that can be used to treat a disease associated with differential expression of the prominin-1 gene, particularly cancer. The method typically includes assaying the ability of the compound to modulate the expression of prominin-1 nucleic acid and thus identifying a compound that can be used to treat a disorder characterized by undesired prominin-1 nucleic acid expression. The assays can be performed in cell-based and cell-free systems. Cell-based assays include cells naturally expressing prominin-1 nucleic acid or recombinant cells genetically engineered to express specific nucleic acid sequences.

[00275] The assay for prominin-1 nucleic acid expression can involve direct assay of nucleic acid levels, such as mRNA levels, or on collateral compounds involved in the signal pathway. Further, the expression of genes that are up- or down-regulated in response to the prominin-1 protein signal pathway can also be assayed. In this embodiment

the regulatory regions of these genes can be operably linked to a reporter gene such as luciferase.

[00276] Thus, modulators of prominin-1 gene expression can be identified in a method wherein a cell is contacted with a candidate compound or agent and the expression of mRNA determined. The level of expression of prominin-1 mRNA in the presence of the candidate compound or agent is compared to the level of expression of prominin-1 mRNA in the absence of the candidate compound or agent. The candidate compound can then be identified as a modulator of nucleic acid expression based on this comparison and be used, for example to treat a disorder characterized by aberrant nucleic acid expression. When expression of mRNA is statistically significantly greater in the presence of the candidate compound than in its absence, the candidate compound is identified as a stimulator of nucleic acid expression. When nucleic acid expression is statistically significantly less in the presence of the candidate compound than in its absence, the candidate compound is identified as an inhibitor of nucleic acid expression.

d. Methods of Monitoring Treatment

[00277] The nucleic acid molecules are also useful for monitoring the effectiveness of modulating compounds or agents on the expression or activity of the prominin-1 gene in clinical trials or in a treatment regimen. Thus, the gene expression pattern can serve as a barometer for the continuing effectiveness of treatment with the compound, particularly with compounds to which a patient can develop resistance. The gene expression pattern can also serve as a marker indicative of a physiological response of the affected cells to the compound. Accordingly, such monitoring would allow either increased administration of the compound or the administration of alternative compounds to which the patient has not become resistant. Similarly, if the level of nucleic acid expression falls below a desirable level, administration of the compound could be commensurately decreased.

e. Treatment Using Nucleic Acids

[00278] The nucleic acid molecules are useful to design antisense constructs to control prominin-1 gene expression in cells, tissues, and organisms. A DNA antisense nucleic acid

molecule is designed to be complementary to a region of the gene involved in transcription, preventing transcription and hence production of prominin-1 protein. An antisense RNA or DNA nucleic acid molecule would hybridize to the mRNA and thus block translation of mRNA into prominin-1 protein.

[00279] The nucleic acid of the present invention may also be used to specifically suppress gene expression by methods such as RNA interference (RNAi), which may also include cosuppression and quelling. This and antisense RNA or DNA of gene suppression are well known in the art. A review of this technique is found in Science 288:1370-1372, 2000. RNAi also operates on a post-transcriptional level and is sequence specific, but suppresses gene expression far more efficiently than antisense RNA. RNAi fragments, particularly double-stranded (ds) RNAi, can be also used to generate loss-of-function phenotypes.

[00280] The present invention relates to isolated RNA molecules (double-stranded; single-stranded) of from about 21 to about 25 nucleotides (nt) which mediate RNAi. As used herein, about 21 to about 25 nt includes nucleotides 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28 and 29 nucleotides in length. The isolated RNAs of the present invention mediate degradation of mRNA, the transcriptional product of a gene. Such mRNA is also referred to herein as mRNA to be degraded. As used herein, the terms RNA, RNA molecule(s), RNA segment(s) and RNA fragment(s) are used interchangeably to refer to RNA that mediates RNA interference. These terms include double-stranded RNA, single-stranded RNA, isolated RNA (partially purified RNA, essentially pure RNA, synthetic RNA, recombinantly produced RNA), as well as altered RNA that differs from naturally occurring RNA by the addition, deletion, substitution and/or alteration of one or more nucleotides. Such alterations can include addition of non-nucleotide material, such as to the end(s) of the 21-25 nt RNA or internally (at one or more nucleotides of the RNA). Nucleotides in the RNA molecules of the present invention can also comprise non-standard nucleotides, including non-naturally occurring nucleotides or deoxyribonucleotides. Collectively, all such altered RNAs are referred to as analogs or analogs of naturally-occurring RNA. RNA of 21-25 nucleotides of the present invention need only be sufficiently similar to natural RNA that it has the ability to mediate RNAi. As used herein

the phrase "mediates RNAi" refers to the ability to distinguish which RNAs are to be degraded by the RNAi machinery or process. RNA that mediates RNAi interacts with the RNAi machinery such that it directs the degradation of particular mRNAs. Such RNA may include RNAs of various structure, including short hairpin RNA.

[00281] In one embodiment, the present invention relates to RNA molecules of about 21 to about 25 nucleotides that direct cleavage of specific mRNA to which their sequence corresponds. It is not necessary that there be perfect correspondence of the sequences, but the correspondence must be sufficient to enable the RNA to direct RNAi cleavage of the target mRNA (Holen et al., Nucleic Acids Res., 33:4704-4710 (2005)). In a particular embodiment, the 21-25 nt RNA molecules of the present invention comprise a 3' hydroxyl group.

[00282] The present invention relates to 21-25 nt RNAs of specific genes, produced by chemical synthesis or recombinant DNA techniques, that mediate RNAi. As used herein, the term isolated RNA includes RNA obtained by any means, including processing or cleavage of dsRNA; production by chemical synthetic methods; and production by recombinant DNA techniques. The invention further relates to uses of the 21-25 nt RNAs, such as for therapeutic or prophylactic treatment and compositions comprising 21-25 nt RNAs that mediate RNAi, such as pharmaceutical compositions comprising 21-25 nt RNAs and an appropriate carrier.

[00283] The present invention also relates to a method of mediating RNA interference of genes of a patient. In one embodiment, RNA of about 21 to about 25 nt which targets the specific mRNA to be degraded is introduced into a patient's cells. The cells are maintained under conditions allowing degradation of the mRNA, resulting in RNA-mediated interference of the mRNA of the gene in the cells of the patient. Treatment of patients with cancer with the RNAi will inhibit the growth and spread of the cancer and reduce the tumor. Treatment of patients using RNAi can also be in combination with other anti-cancer compounds. The RNAi may be used in combination with other treatment modalities, such as chemotherapy, cryotherapy, hyperthermia, radiation therapy, and other similar

treatments. In one embodiment, a chemotherapy agent was combined with the RNAi. In another embodiment, a chemotherapy named GEMZAR (gemcitabine HCl) was used.

[00284] Treatment of cancer or tumors in patients requires introduction of the RNA into the cancer or tumor cells. RNA may be directly introduced into the cell, or introduced extracellularly into a cavity, interstitial space, into the circulation of a patient, or introduced orally. Methods for oral introduction include direct mixing of the RNA with food, as well as engineered approaches in which a species that is used as food is engineered to express the RNA and then ingested. Physical methods of introducing nucleic acids, for example, injection directly into the cell or extracellular injection into the patient, may also be used. Vascular or extravascular circulation, the blood or lymph system, and the cerebrospinal fluid are sites where the RNA may be introduced. RNA may be introduced into an embryonic stem cell, or another multipotent cell derived from the patient. Physical methods of introducing nucleic acids include injection of a solution containing the RNA, bombardment by particles covered by the RNA, soaking cells or tissue in a solution of the RNA, or electroporation of cell membranes in the presence of the RNA. A viral construct packaged into a viral particle may be used to introduce an expression construct into the cell, with the construct expressing RNA. Other methods known in the art for introducing nucleic acids to cells may be used, such as lipid-mediated carrier transport, chemical-mediated transport, and the like. Thus the RNA may be introduced along with components that perform one or more of the following activities: enhance RNA uptake by the cell, promote annealing of the duplex strands, stabilize the annealed strands, or otherwise increase inhibition of the target gene. The RNAi may be used in combination with other treatment modalities, such as chemotherapy, cryotherapy, hyperthermia, radiation therapy, and the like.

[00285] The present invention may be used alone or as a component of a kit having at least one of the reagents necessary to carry out the in vitro or in vivo introduction of RNA to tissue or patients. Preferred components are the dsRNA and a vehicle that promotes introduction of the dsRNA. Such a kit may also include instructions to allow a user of the kit to practice the invention.

[00286] Alternatively, a class of antisense molecules can be used to inactivate mRNA in order to decrease expression of prominin-1 nucleic acid. Accordingly, these molecules can treat a disorder characterized by abnormal or undesired prominin-1 nucleic acid expression. This technique involves cleavage by means of ribozymes containing nucleotide sequences complementary to one or more regions in the mRNA that attenuate the ability of the mRNA to be translated. Possible regions include coding regions and particularly coding regions corresponding to the catalytic and other functional activities of the prominin-1 protein, such as substrate binding.

[00287] The nucleic acid molecules can be used for gene therapy in patients containing cells that are aberrant in prominin-1 gene expression. Thus, recombinant cells, which include the patient's cells that have been engineered ex vivo and returned to the patient, are introduced into an individual where the cells produce the desired prominin-1 protein to treat the individual.

9. Diagnosis Using Prominin-1 Protein

Protein Detection

[00288] The present invention provides methods for diagnosing or detecting the differential presence of prominin-1 protein. Where prominin-1 is overexpressed in diseased cells, prominin-1 protein is detected directly.

[00289] The information obtained is also used to determine prognosis and appropriate course of treatment. For example, it is contemplated that individuals with a specific prominin-1 expression or stage of disease may respond differently to a given treatment that individuals lacking prominin-1 expression. The information obtained from the diagnostic methods of the present invention thus provides for the personalization of diagnosis and treatment.

[00290] In one embodiment, the present invention provides a method for monitoring disease treatment in a subject comprising: determining the level of prominin-1 protein or any fragment(s) or peptide(s) thereof in a test sample from said subject, wherein a level of

said prominin-1 protein similar to the level of said protein in a test sample from a healthy subject, or the level established for a healthy subject, is indicative of successful treatment.

[00291] In another embodiment, the present invention provides a method for diagnosing recurrence of disease following successful treatment in a subject comprising: determining the level of prominin-1 protein or any fragment(s) or peptide(s)thereof in a test sample from said subject; wherein a changed level of said prominin-1 protein relative to the level of said protein in a test sample from a healthy subject, or the level established for a healthy subject, is indicative of recurrence of diseases.

[00292] In yet another embodiment, the present invention provides a method for diagnosing or detecting disease in a subject comprising: determining the level of prominin-1 protein or any fragment or peptides thereof in a test sample from said subject; wherein a differential level of said prominin-1 protein relative to the level of said protein in a test sample from a healthy subject, or the level established for a healthy subject, is indicative of disease.

[00293] These methods are also useful for diagnosing diseases that show differential protein expression. As describe earlier, normal, control or standard values or level established from a healthy subject for protein expression are established by combining body fluids or tissue, cell extracts taken from a normal healthy mammalian or human subject with specific antibodies to a protein under conditions for complex formation. Standard values for complex formation in normal and diseased tissues are established by various methods, often photometric means. Then complex formation as it is expressed in a subject sample is compared with the standard values. Deviation from the normal standard and toward the diseased standard provides parameters for disease diagnosis or prognosis while deviation away from the diseased and toward the normal standard may be used to evaluate treatment efficacy.

[00294] In yet another embodiment, the present invention provides a detection or diagnostic method of prominin-1 by using LC/MS. The proteins from cells can be prepared by methods known in the art (for example, Zhang et al., *Nature Biotechnology* 21(6):660-666 (2003)). The differential expression of proteins in disease and healthy samples are

quantitated using Mass Spectrometry and ICAT (Isotope Coded Affinity Tag) labeling, which is known in the art. ICAT is an isotope label technique that allows for discrimination between two populations of proteins, such as a healthy and a disease sample. The LC/MS spectra are collected for the labeled samples. The raw scans from the LC/MS instrument are subjected to peak detection and noise reduction software. Filtered peak lists are then used to detect 'features' corresponding to specific peptides from the original sample(s). Features are characterized by their mass/charge, charge, retention time, isotope pattern and intensity.

[00295] The intensity of a peptide present in both healthy and disease samples can be used to calculate the differential expression, or relative abundance, of the peptide. The intensity of a peptide found exclusively in one sample can be used to calculate a theoretical expression ratio for that peptide (singleton). Expression ratios are calculated for each peptide of each replicate of the experiment. Thus overexpression or under expression of prominin-1 protein or peptide are similar to the expression pattern in a test subject indicates the likelihood of having a disease, particularly cancer, or an associated pathology.

[00296] Immunological methods for detecting and measuring complex formation as a measure of protein expression using either specific polyclonal or monoclonal antibodies are known in the art. Examples of such techniques include enzyme-linked immunosorbent assays (ELISAs), radioimmunoassays (RIAs), fluorescence-activated cell sorting (FACS) and antibody arrays. Such immunoassays typically involve the measurement of complex formation between the protein and its specific antibody. These assays and their quantitation against purified, labeled standards are well known in the art (Ausubel, supra, unit 10.1-10.6). A two-site, monoclonal-based immunoassay utilizing antibodies reactive to two non-interfering epitopes is preferred, but a competitive binding assay may be employed (Pound (1998) Immunochemical Protocols, Humana Press, Totowa N.J.). More immunological detections are described in section below.

[00297] For diagnostic applications, the antibody or its variant typically will be labeled with a detectable moiety. Numerous labels are available which can be generally grouped into the following categories:

[00298] (a) Radioisotopes, such as ^{36}S, ^{14}C, ^{125}I, ^{3}H, and ^{131}I. The antibody variant can be labeled with the radioisotope using the techniques described in Current Protocols in Immunology, vol 1-2, Coligen et al., Ed., Wiley-Interscience, New York, Pubs. (1991-2006) for example and radioactivity can be measured using scintillation counting.

[00299] (b) Fluorescent labels such as rare earth chelates (europium chelates) or fluorescein and its derivatives, rhodamine and its derivatives, dansyl, Lissamine, phycoerythrin and Texas Red are available. The fluorescent labels can be conjugated to the antibody variant using the techniques disclosed in Current Protocols in Immunology, supra, for example. Fluorescence can be quantified using a fluorometer.

[00300] (c) Various enzyme-substrate labels are available and U.S. Pat. Nos. 4,275,149 and 4,318,980 provide a review of some of these. The enzyme generally catalyzes a chemical alteration of the chromogenic substrate which can be measured using various techniques. For example, the enzyme may catalyze a color change in a substrate, which can be measured spectrophotometrically. Alternatively, the enzyme may alter the fluorescence or chemiluminescence of the substrate. Techniques for quantifying a change in fluorescence are described above. The chemiluminescent substrate becomes electronically excited by a chemical reaction and may then emit light which can be measured (using a chemiluminometer, for example) or donates energy to a fluorescent acceptor. Examples of enzymatic labels include luciferases (e.g., firefly luciferase and bacterial luciferase; U.S. Pat. No. 4,737,456),. luciferin, 2,3-dihydrophthalazinediones, malate dehydrogenase, urease, peroxidase such as horseradish peroxidase (HRPO), alkaline phosphatase, β-galactosidase, glucoamylase, lysozyme, saccharide oxidases (e.g., glucose oxidase, galactose oxidase, and glucose-6-phosphate dehydrogenase), heterocyclic oxidases (such as uricase and xanthine oxidase), lactoperoxidase, microperoxidase, and the like. Techniques for conjugating enzymes to antibodies are described in O'Sullivan et al., Methods for the Preparation of Enzyme-Antibody Conjugates for Use in Enzyme Immunoassay, in Methods in Enzyme. (Ed. J. Langone & H. Van Vunakis), Academic press, New York, 73: 147-166 (1981), and are also described in Section III.

[00301] Sometimes, the label is indirectly conjugated with the antibody. The skilled artisan will be aware of various techniques for achieving this. For example, the antibody can be conjugated with biotin and any of the three broad categories of labels mentioned above can be conjugated with avidin, or vice versa. Biotin binds selectively to avidin and thus, the label can be conjugated with the antibody in this indirect manner. Alternatively, to achieve indirect conjugation of the label with the antibody, the antibody is conjugated with a small hapten (e.g., digoxin) and one of the different types of labels mentioned above is conjugated with an anti-hapten antibody (e.g., anti-digoxin antibody). Thus, indirect conjugation of the label with the antibody can be achieved.

[00302] The biological samples can then be tested directly for the presence of prominin-1 by assays (e.g., ELISA or radioimmunoassay) and format (e.g., microwells, dipstick, etc., as described in International Patent Publication WO 93/03367). Alternatively, proteins in the sample can be size separated (e.g., by polyacrylamide gel electrophoresis (PAGE)), in the presence or absence of sodium dodecyl sulfate (SDS), and the presence of prominin-1 detected by immunoblotting (e.g., Western blotting). Immunoblotting techniques are generally more effective with antibodies generated against a peptide corresponding to an epitope of a protein, and hence, are particularly suited to the present invention.

[00303] Antibody binding may be detected also by "sandwich" immunoassays, immunoradiometric assays, gel diffusion precipitation reactions, immunodiffusion assays, in situ immunoassays (e.g., using colloidal gold, enzyme or radioisotope labels, for example), precipitation reactions, agglutination assays (e.g., gel agglutination assays, hemagglutination assays, etc.), complement fixation assays, immunofluorescence assays, protein A assays, and immunoelectrophoresis assays, etc.

[00304] In one embodiment, antibody binding is detected by detecting a label on the primary antibody. In another embodiment, the primary antibody is detected by detecting binding of a secondary antibody or reagent to the primary antibody. In a further embodiment, the secondary antibody is labeled. Many means are known in the art for detecting binding in an immunoassay and are within the scope of the present invention. As is well known in the art, the immunogenic peptide should be provided free of the carrier

molecule used in any immunization protocol. For example, if the peptide is conjugated to KLH, it may be conjugated to BSA, or used directly, in a screening assay. In some embodiments, an automated detection assay is utilized. Methods for the automation of immunoassays are well known in the art (see e.g., U.S. Pat. Nos. 5,885,530: 4,981,785: 6,159,750: and 5,358,691, each of which is herein incorporated by reference). In some embodiments, the analysis and presentation of results is also automated. For example, in some embodiments, software that generates a prognosis based on the presence or absence of a series of antigens is utilized.

[00305] Competitive binding assays rely on the ability of a labeled standard to compete with the test sample for binding with a limited amount of antibody. The amount of antigen in the test sample is inversely proportional to the amount of standard that becomes bound to the antibodies. To facilitate determining the amount of standard that becomes bound, the antibodies generally are insolubilized before or after the competition. As a result, the standard and test sample that are bound to the antibodies may conveniently be separated from the standard and test sample, which remain unbound.

[00306] Sandwich assays involve the use of two antibodies, each capable of binding to a different immunogenic portion, or epitope, or the protein to be detected. In a sandwich assay, the test sample to be analyzed is bound by a first antibody, which is immobilized on a solid support, and thereafter a second antibody binds to the test sample, thus forming an insoluble three-part complex. See e.g., U.S. Pat. No. 4,376,110. The second antibody may itself be labeled with a detectable moiety (direct sandwich assays) or may be measured using an anti-immunoglobulin antibody that is labeled with a detectable moiety (indirect sandwich assay). For example, one type of sandwich assay is an ELISA assay, in which case the detectable moiety is an enzyme.

[00307] The antibodies may also be used for in vivo diagnostic assays. Generally, the antibody is labeled with a radionucleotide (such as ^{111}In, ^{99}Tc, ^{14}C, ^{131}I, ^{3}H, ^{32}P or ^{35}S) so that the tumor can be localized using immunoscintiography. In one embodiment, antibodies or fragaments thereof bind to the extracellular domains of two or more prominin-1 targets and the affinity value (Kd) is less than 1×10^8 M.

[00308] Antibodies for diagnostic use may be labeled with probes suitable for detection by various imaging methods. Methods for detection of probes include, but are not limited to, fluorescence, light, confocal and electron microscopy; magnetic resonance imaging and spectroscopy; fluoroscopy, computed tomography and positron emission tomography. Suitable probes include, but are not limited to, fluorescein, rhodamine, eosin and other fluorophores, radioisotopes, gold, gadolinium and other lanthanides, paramagnetic iron, fluorine-18 and other positron-emitting radionuclides. Additionally, probes may be bi- or multi-functional and be detectable by more than one of the methods listed. These antibodies may be directly or indirectly labeled with said probes. Attachment of probes to the antibodies includes covalent attachment of the probe, incorporation of the probe into the antibody, and the covalent attachment of a chelating compound for binding of probe, amongst others well recognized in the art.

[00309] For immunohistochemistry, the disease tissue sample may be fresh or frozen or may be embedded in paraffin and fixed with a preservative such as formalin (see Example). The fixed or embedded section contains the sample are contacted with a labeled primary antibody and secondary antibody, wherein the antibody is used to detect prominin-1 protein expression in situ. The detailed procedure is shown in the Examples.

[00310] Antibodies against prominin-1 protein or peptides are useful to detect the presence of one of the proteins of the present invention in cells or tissues to determine the pattern of expression of the protein among various tissues in an organism and over the course of normal development.

[00311] Further, such antibodies can be used to detect protein in situ, in vitro, or in a cell lysate or supernatant in order to evaluate the abundance and pattern of expression. Also, such antibodies can be used to assess abnormal tissue distribution or abnormal expression during development or progression of a biological condition. Antibody detection of circulating fragments of the full length protein can be used to identify turnover.

[00312] Further, the antibodies can be used to assess expression in disease states such as in active stages of the disease or in an individual with a predisposition toward disease related to the protein's function. When a disorder is caused by an inappropriate tissue

distribution, developmental expression, level of expression of the protein, or expressed/processed form, the antibody can be prepared against the normal protein. If a disorder is characterized by a specific mutation in the protein, antibodies specific for this mutant protein can be used to assay for the presence of the specific mutant protein.

[00313] The antibodies can also be used to assess normal and aberrant subcellular localization of cells in the various tissues in an organism. The diagnostic uses can be applied, not only in genetic testing, but also in monitoring a treatment modality. Accordingly, where treatment is ultimately aimed at correcting expression level or the presence of aberrant sequence and aberrant tissue distribution or developmental expression, antibodies directed against the protein or relevant fragments can be used to monitor therapeutic efficacy. More detection and diagnostic methods are described in detail below.

[00314] Additionally, antibodies are useful in pharmacogenomic analysis. Thus, antibodies prepared against polymorphic proteins can be used to identify individuals that require modified treatment modalities. The antibodies are also useful as diagnostic tools, as an immunological marker for aberrant protein analyzed by electrophoretic mobility, isoelectric point, tryptic peptide digest, and other physical assays known to those in the art.

[00315] The antibodies are also useful for tissue typing. Where a specific protein has been correlated with expression in a specific tissue, antibodies that are specific for this protein can be used to identify a tissue type.

[00316] The invention also encompasses kits for using antibodies to detect the presence of a protein in a biological sample. The kit can comprise antibodies such as a labeled or labelable antibody and a compound or agent for detecting protein in a biological sample; means for determining the amount of protein in the sample; means for comparing the amount of protein in the sample with a standard; and instructions for use. Such a kit can be supplied to detect a single protein or epitope or can be configured to detect one of a multitude of epitopes, such as in an antibody detection array. Arrays are described in detail below for nucleic acid arrays and similar methods have been developed for antibody arrays.

10. Methods of Treatment Based on Prominin-1 Protein

a. Antibody Therapy

[00317] The antibodies of the present invention can be used for therapeutic purposes. It is contemplated that the antibodies of the present invention may be used to treat a mammal, preferably a human, with a disease. The antibodies can be delivered alone or conjugated to one or more therapeutic agents.

[00318] Antibodies can also also be useful for modulating (agonizing or antagonizing) protein function, and may be applied in a therapeutic context in which treatment involves modulating the protein's function. Antibodies can be prepared against, for example, specific portions of a protein that contain domains required for protein function, or against intact protein that is associated with a cell membrane.

[00319] The antibodies of the present invention can also be used for enhancing the immune response. The antibodies can be administered in amounts similar to those used for other therapeutic administrations of antibodies. For example, pooled gamma globulin can be administered at a range of about 1 mg to about 100 mg per patient.

[00320] Antibodies reactive with prominin-1 proteins can be administered alone or in conjunction with other therapies, such as anti-cancer therapies, to a mammal afflicted with cancer or other disease. Examples of anti-cancer therapies include, but are not limited to, chemotherapy, radiation therapy, and adoptive immunotherapy therapy with TIL (Tumor Infiltration Lymphocytes).

[00321] The selection of an antibody subclass for therapy will depend upon the nature of the antigen to be acted upon. For example, an IgM may be preferred in situations where the antigen is highly specific for the diseased target and rarely occurs on normal cells. However, where the disease-associated antigen is also expressed in normal tissues, although at much lower levels, the IgG subclass may be preferred, since the binding of at least two IgG molecules in close proximity is required to activate complement, less complement mediated damage may occur in the normal tissues which express smaller amounts of the antigen and, therefore, bind fewer IgG antibody molecules. Furthermore,

IgG molecules by being smaller may be more able than IgM molecules to localize to the diseased tissue.

[00322] The mechanism for antibody therapy is that the therapeutic antibody recognizes a cell surface protein or a cytosolic protein that is expressed or preferably, overexpressed in a diseased cell. By NK cell or complement activation, or conjugation of the antibody with an immunotoxin or radiolabel, the interaction can abrogate ligand/receptor interaction or activation of apoptosis.

[00323] The potential mechanisms of antibody-mediated cytotoxicity of diseased cells are phagocyte (antibody dependent cellular cytotoxicity (ADCC)) (see Example), complement (Complement-mediated cytotoxicity (CMC)) (see Example), naked antibody (receptor cross-linking apoptosis and growth factor inhibition), or targeted payload labeled with a therapeutic agent, such as a radionuclide, immunotoxin or immunochemotherapeutic or other therapeutic agent.

[00324] In one embodiment, the antibody is administered to a nonhuman mammal for the purposes of obtaining preclinical data, for example. Exemplary nonhuman mammals to be treated include nonhuman primates, dogs, cats, rodents and other mammals in which preclinical studies are performed. Such mammals may be established animal models for a disease to be treated with the antibody or may be used to study toxicity of the antibody of interest. In each of these embodiments, dose escalation studies may be performed on the mammal.

[00325] The antibody is administered by any suitable means, including parenteral, subcutaneous, intraperitoneal, intrapulmonary, and intranasal, and, if desired for local immunomodulatory treatment, intralesional administration. Parenteral infusions include intramuscular, intravenous, intraarterial, intraperitoneal, or subcutaneous administration. In addition, the antibody variant is suitably administered by pulse infusion, particularly with declining doses of the antibody variant. Preferably the dosing is given by injections, most preferably intravenous or subcutaneous injections, depending in part on whether the administration is brief or chronic.

[00326] For the prevention or treatment of a disease, the appropriate dosage of the antibody will depend on the type of disease to be treated, the severity and the course of the disease, whether the antibody is administered for preventive or therapeutic purposes, previous therapy, the patient's clinical history and response to the antibody, and the discretion of the attending physician.

[00327] Depending on the type and severity of the disease, about 1 μg/kg to 150 mg/kg (e.g., 0.1-20 mg/kg) of antibody is an initial candidate dosage for administration to the patient, whether, for example, by one or more separate administrations, or by continuous infusion. A typical daily dosage might range from about 1 μg/kg to 100 mg/kg or more, depending on the factors mentioned above. For repeated administrations over several days or longer, depending on the condition, the treatment is sustained until a desired suppression of disease symptoms occurs. However, other dosage regimens may be useful. The progress of this therapy is easily monitored by conventional techniques and assays.

[00328] An antibody drug conjugate can be administered from about 1 μg/kg to 50 mg/kg, typically from about 0.1-20 mg/kg, whether, for example, by one or more separate administrations, or by continuous infusion. A typical daily dosage might range from about 0.1 mg/kg to 10 mg/kg, from about 0.3 mg/kg to about 7.5 mg/kg, depending on the factors mentioned above. For repeated administrations over several days or longer, depending on the condition, the treatment is sustained until a desired suppression of disease symptoms occurs. However, other dosage regimens may be useful. The progress of this therapy is easily monitored by conventional techniques and assays.

[00329] The antibody composition will be formulated, dosed and administered in a manner consistent with good medical practice. Factors for consideration in this context include the particular disorder being treated, the particular mammal being treated, the clinical condition of the individual patient, the cause of the disorder, the site of delivery of the agent, the method of administration, the scheduling of administration, and other factors known to medical practitioners.

[00330] The therapeutically effective amount of the antibody to be administered will be governed by such considerations, and is the minimum amount necessary to prevent,

ameliorate, or treat a disease or disorder. The antibody may optionally be formulated with one or more therapeutic agents currently used to prevent or treat the disorder in question. For example, an antibody can be administered as a co-therapy with a standard of care therapeutic for the specific disease being treated.

[00331] Suitable therapeutic agents in this regard include radionuclides, differentiation inducers, chemotherapeutic drugs, toxins, and derivatives thereof. Exemplary radionuclides include ^{90}Y, ^{123}I, ^{125}I, ^{131}I, ^{186}Re, ^{188}Re ^{211}At, and ^{212}Bi. Exemplary chemotherapeutic drugs include methotrexate, vindesine, adriamycin, taxol, cisplatinum, irinotecan, leucovorin, and pyrimidine and purine analogs (e.g., 5-fluorouracil). Exemplary differentiation inducers include phorbol esters and butyric acid. Exemplary toxins include ricin, abrin, diptheria toxin, cholera toxin, gelonin, *Pseudomonas* exotoxin, *Shigella* toxin, and pokeweed antiviral protein.

b. *Other Immunotherapy*

[00332] Peptides derived from the prominin-1 protein sequence may be modified to increase their immunogenicity by enhancing the binding of the peptide to the MHC molecules in which the peptide is presented. The peptide or modified peptide may be conjugated to a carrier molecule to enhance the antigenicity of the peptide. Examples of carrier molecules, include, but are not limited to, human albumin, bovine albumin, lipoprotein and keyhole limpet hemo-cyanin ("Basic and Clinical Immunology" (1991) Stites, D. P. and Terr A. I. (eds) Appleton and Lange, Norwalk Conn., San Mateo, Calif.).

[00333] An "immunogenic peptide" is a peptide, which comprises an allele-specific motif such that the peptide will bind the MHC allele (HLA in human) and be capable of inducing a CTL (cytotoxic T-lymphocytes) response. Thus, immunogenic peptides are capable of binding to an appropriate class I or II MHC molecule and inducing a cytotoxic T cell or T helper cell response against the antigen from which the immunogenic peptide is derived.

[00334] Alternatively, amino acid sequence variants of the peptide can be prepared by altering the nucleic acid sequence of the DNA which encodes the peptide, or by peptide synthesis. At the genetic level, these variants ordinarily are prepared by site-directed

mutagenesis of nucleotides in the DNA encoding the peptide molecule, thereby producing DNA encoding the variant, and thereafter expressing the DNA in recombinant cell culture. The variants typically exhibit the same qualitative biological activity as the nonvariant peptide.

[00335] The recombinant or natural protein, peptides, or fragment thereof of prominin-1, or modified peptides, may be used as a vaccine either prophylactically or therapeutically. When provided prophylactically the vaccine is provided in advance of any evidence of disease, particularly, cancer. The prophylactic administration of the disease vaccine should serve to prevent or attenuate diseases, preferably cancer, in a mammal.

[00336] Preparation of vaccine uses recombinant protein or peptide expression vectors comprising a nucleic acid sequence encoding all or part of the prominin-1 protein. Examples of vectors that may be used in the aforementioned vaccines include, but are not limited to, defective retroviral vectors, adenoviral vectors vaccinia viral vectors, fowl pox viral vectors, or other viral vectors (Mulligan, R. C., (1993) Science 260:926-932). The vectors can be introduced into a mammal either prior to any evidence of the disease or to mediate regression of the disease in a mammal afflicted with disease. Examples of methods for administering the viral vector into the mammals include, but are not limited to, exposure of cells to the virus *ex vivo*, or injection of the retrovirus or a producer cell line of the virus into the affected tissue or intravenous administration of the virus. Alternatively the vector may be administered locally by direct injection into the cancer lesion or topical application in a pharmaceutically acceptable carrier. The quantity of viral vector, carrying all or part of the prominin-1 nucleic acid sequence, to be administered is based on the titer of virus particles. A preferred range may be about 10^6 to about 10^{11} virus particles per mammal, preferably a human.

[00337] After immunization the efficacy of the vaccine can be assessed by the production of antibodies or immune cells that recognize the antigen, as assessed by specific lytic activity or specific cytokine production or by tumor regression. One skilled in the art would know the conventional methods to assess the aforementioned parameters. If the mammal to be immunized is already afflicted with cancer, the vaccine can be administered in

conjunction with other therapeutic treatments. Examples of other therapeutic treatments includes, but are not limited to, adoptive T cell immunotherapy, coadministration of cytokines or other therapeutic drugs for cancer.

[00338] Alternatively all or parts thereof of a substantially or partially purified the prominin-1 protein or their peptides may be administered as a vaccine in a pharmaceutically acceptable carrier. Ranges of the protein that may be administered are about 0.001 to about 100 mg per patient, preferred doses are about 0.01 to about 100 mg per patient. Immunization may be repeated as necessary, until a sufficient titer of anti-immunogen antibody or immune cells has been obtained.

[00339] In yet another alternative embodiment a viral vector, such as a retroviral vector, can be introduced into mammalian cells. Examples of mammalian cells into which the retroviral vector can be introduced include, but are not limited to, primary mammalian cultures or continuous mammalian cultures, COS cells, NIH3T3, or 293 cells (ATTC #CRL 1573), dendritic cells. The means by which the vector carrying the gene may be introduced into a cell includes, but is not limited to, microinjection, electroporation, transfection or transfection using DEAE dextran, lipofection, calcium phosphate or other procedures known to one skilled in the art (Sambrook et al. 3rd. ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, (2001).

[00340] The vaccine formulation of the present invention comprises an immunogen that induces an immune response directed against the cancer associated antigen such as prominin-1 protein, and in nonhuman primates and finally in humans. The safety of the immunization procedures is determined by looking for the effect of immunization on the general health of the immunized animal (weight change, fever, appetite behavior etc.) and looking for pathological changes on autopsies. After initial testing in animals, cancer patients can be tested. Conventional methods would be used to evaluate the immune response of the patient to determine the efficiency of the vaccine.

[00341] In one embodiment mammals, preferably human, at high risk for disease, particularly cancer, are prophylactically treated with the vaccines of this invention. Examples include, but are not limited to, humans with a family history of a disease,

humans with a history of disease, particular cancer, or humans afflicted with a disease, such as cancer that has been previously resected and therefore at risk for reoccurrence. When provided therapeutically, the vaccine is provided to enhance the patient's own immune response to the disease antigen present on the disease cells or present during advanced stage of the disease. The vaccine, which acts as an immunogen, may be a cell, cell lysate from cells transfected with a recombinant expression vector, or a culture supernatant containing the expressed protein. Alternatively, the immunogen is a partially or substantially purified recombinant protein, peptide or analog thereof or modified peptides or analogs thereof. The proteins or peptides may be conjugated with lipoprotein or administered in liposomal form or with adjuvant.

[00342] While it is possible for the immunogen to be administered in a pure or substantially pure form, it is preferable to present it as a pharmaceutical composition, formulation or preparation, as discussed hereinabove.

[00343] Vaccination can be conducted by conventional methods. For example, the immunogen can be used in a suitable diluent such as saline or water, or complete or incomplete adjuvants. Further, the immunogen may or may not be bound to a carrier to make the protein immunogenic. Examples of such carrier molecules include but are not limited to bovine serum albumin (BSA), keyhole limpet hemocyanin (KLH), tetanus toxoid, and the like. The immunogen also may be coupled with lipoproteins or administered in liposomal form or with adjuvants. The immunogen can be administered by any route-appropriate for antibody production such as intravenous, intraperitoneal, intramuscular, subcutaneous, and the like. The immunogen may be administered once or at periodic intervals until a significant titer of anti-prominin-1 immune cells or anti-prominin-1 antibody is produced. The presence of anti-prominin-1 immune cells may be assessed by measuring the frequency of precursor CTL (cytotoxic T-lymphocytes) against prominin-1 antigen prior to and after immunization by a CTL precursor analysis assay (Coulie, P. et al., (1992) International Journal Of Cancer 50:289-297). The antibody may be detected in the serum using the immunoassay described above.

[00344] The safety of the immunization procedures is determined by examining the effect of immunization on the general health of the immunized animal (fever, change in weight, appetite, behavior etc.) and pathological changes on autopsies. After initial testing in animals, human patients can be tested. Conventional methods would be used to evaluate the immune response of the patient to determine the efficiency of the vaccine.

[00345] In yet another embodiment of this invention all, part, or parts of the prominin-1 protein or peptides or fragments thereof, or modified peptides, may be exposed to dendritic cells cultured *in vitro*. The cultured dendritic cells provide a means of producing T-cell dependent antigens comprised of dendritic cell modified antigen or dendritic cells pulsed with antigen, in which the antigen is processed and expressed on the antigen activated dendritic cell. The prominin-1 antigen activated dendritic cells or processed dendritic cell antigens may be used as immunogens for vaccines or for the treatment of diseases, particularly cancer. The dendritic cells should be exposed to the antigen for sufficient time to allow the antigens to be internalized and presented on the dendritic cells surface. The resulting dendritic cells or the dendritic-cell processed antigens can then be administered to an individual in need of therapy. Such methods are described in Steinman et al. (WO93/208185) and in Banchereau et al. (EPO Application 0563485A1).

[00346] In yet another aspect of this invention T-cells isolated from individuals can be exposed to prominin-1 protein, peptides or fragment thereof, or modified peptides in vitro and then administered to a patient in need of such treatment in a therapeutically effective amount. Examples of where T-lymphocytes can be isolated include but are not limited to, peripheral blood cells lymphocytes (PBL), lymph nodes, or tumor infiltrating lymphocytes (TIL). Such lymphocytes can be isolated from the individual to be treated or from a donor by methods known in the art and cultured in vitro (Kawakami, Y. et al. (1989) J. Immunol. 142: 2453-3461). Lymphocytes are cultured in media such as RPMI or RPMI 1640 or AIM V for 1-10 weeks. Viability is assessed by trypan blue dye exclusion assay. Examples of how these sensitized T-cells can be administered to the mammal include but are not limited to, intravenously, intraperitoneally or intralesionally. Parameters that may be assessed to determine the efficacy of these sensitized T-lymphocytes include, but are not limited to, production of immune cells in the mammal being treated or tumor regression. Conventional

methods are used to assess these parameters. Such treatment can be given in conjunction with cytokines or gene modified cells (Rosenberg, S. A. et al. (1992) Human Gene Therapy, 3: 75-90; Rosenberg, S. A. et al. (1992) Human Gene Therapy, 3: 57-73).

[00347] The present invention is further described by the following examples, which are provided solely to illustrate the invention by reference to specific embodiments. This exemplification, while illustrating certain aspects of the invention, does not offer the limitations or circumscribe the scope of the disclosed invention.

11. Screening Methods Using Proteins

[00348] The prominin-1 protein and polypeptide can be used to identify compounds or agents that modulate prominin-1 activity of the protein in its natural state or an altered form that causes a specific disease or pathology associated with prominin-1. Both prominin-1 of the present invention and appropriate variants and fragments can be used in high-throughput screens to assay candidate compounds for the ability to bind to prominin-1. These compounds can be further screened against functional prominin-1 to determine the effect of the compound on prominin-1 activity. Further, these compounds can be tested in animal or invertebrate systems to determine activity/effectiveness. Compounds can be identified that activate (agonist) or inactivate (antagonist) prominin-1 to a desired degree.

[00349] Both prominin-1 of the present invention and appropriate variants and fragments can be used in high-throughput screening to assay candidate compounds for the ability to bind to prominin-1. These compounds can be further screened against functional prominin-1 to determine the effect of the compound on prominin-1 activity. Further, these compounds can be tested in animal or invertebrate systems to determine activity/effectiveness. Compounds can be identified that activate (agonist) or inactivate (antagonist) prominin-1 to a desired degree.

[00350] Further, the proteins of the present invention can be used to screen a compound or an agent for the ability to stimulate or inhibit interaction between prominin-1 protein and a molecule that normally interacts with prominin-1 protein, e.g. a substrate or an extracellular binding ligand or a component of the signal pathway that prominin-1 protein normally

interacts (for example, a cytosolic signal protein). Such assays typically include the steps of combining prominin-1 protein with a candidate compound under conditions that allow prominin-1 protein, or fragment, to interact with the target molecule, and to detect the formation of a complex between the protein and the target or to detect the biochemical consequence of the interaction with prominin-1 protein and the target, such as any of the associated effects of signal transduction such as protein phosphorylation, cAMP turnover, and adenylate cyclase activation, etc.

[00351] Candidate compounds or agents include 1) peptides such as soluble peptides, including Ig-tailed fusion peptides and members of random peptide libraries (see, e.g., Lam et al., Nature 354:82-84 (1991); Houghten et al., Nature 354:84-86 (1991)) and combinatorial chemistry-derived molecular libraries made of D- and/or L- configuration amino acids; 2) phosphopeptides (e.g., members of random and partially degenerate, directed phosphopeptide libraries, see, e.g., Songyang et al., Cell 72:767-778 (1993)); 3) antibodies (e.g., polyclonal, monoclonal, humanized, anti-idiotypic, chimeric, and single chain antibodies as well as Fab, F(ab')2, Fab expression library fragments, and epitope-binding fragments of antibodies); and 4) small organic and inorganic molecules (e.g., molecules obtained from combinatorial and natural product libraries).

[00352] One candidate compound or agent is a soluble fragment of prominin-1 that competes for substrate binding. Other candidate compounds include mutant prominin-1 or appropriate fragments containing mutations that affect prominin-1 function and thus compete for substrate. Accordingly, a fragment that competes for substrate, for example with a higher affinity, or a fragment that binds substrate but does not allow release, is encompassed by the invention.

[00353] Any of the biological or biochemical functions mediated by prominin-1 can be used as an endpoint assay to identify an agent that modulates prominin-1 activity. These include all of the biochemical or biochemical/biological events described herein, in the references cited herein, incorporated by reference for these endpoint assay targets, and other functions known to those of ordinary skill in the art or that can be readily identified.

Specifically, a biological function of a cell or tissues that expresses prominin-1 can be assayed.

[00354] A substrate-binding region can be used that interacts with a different substrate than one which is recognized by the native prominin-1. Accordingly, a different set of signal transduction components is available as an end-point assay for activation. This allows for assays to be performed in other than the specific host cell from which prominin-1 is derived.

[00355] Competition binding assays may also be used to discover compounds that interact with prominin-1 (e.g., binding partners and/or ligands). Thus, a compound is exposed to prominin-1 polypeptide under conditions that allow the compound to bind or to otherwise interact with the polypeptide. Soluble prominin-1 polypeptide is also added to the mixture. If the test compound interacts with the soluble prominin-1 polypeptide, it decreases the amount of complex formed or activity from prominin-1. This type of assay is particularly useful in cases in which compounds are sought that interact with specific regions of prominin-1. Thus, the soluble polypeptide that competes with the target prominin-1 region is designed to contain peptide sequences corresponding to the region of interest.

[00356] To perform cell free drug screening assays, it is sometimes desirable to immobilize either the prominin-1 protein, or fragment, or its target molecule to facilitate separation of complexes from uncomplexed forms of one or both of the proteins, as well as to accommodate automation of the assay.

[00357] Techniques for immobilizing proteins on matrices can be used in the drug screening assays. In one embodiment, a fusion protein can be provided which adds a domain that allows the protein to be bound to a matrix. For example, glutathione-S-transferase fusion proteins can be adsorbed onto glutathione SEPHAROSE beads (Sigma Chemical, St. Louis, MO) or glutathione derivatized microtitre plates, which are then combined with the cell lysates (e.g., ^{35}S-labeled) and the candidate compound, and the mixture incubated under conditions conducive to complex formation (e.g., at physiological conditions for salt and pH). Following incubation, the beads are washed to remove any unbound label, and the matrix immobilized and radiolabel determined directly, or in the

supernatant after the complexes are dissociated. Alternatively, the complexes can be dissociated from the matrix, separated by SDS-PAGE, and the level of prominin-1-binding protein found in the bead fraction quantitated from the gel using standard electrophoretic techniques. For example, either the polypeptide or its target molecule can be immobilized utilizing conjugation of biotin and streptavidin using techniques well known in the art. Alternatively, antibodies reactive with the protein but which do not interfere with binding of the protein to its target molecule can be derivatized to the wells of the plate, and the protein trapped in the wells by antibody conjugation. Preparations of prominin-1-binding protein and a candidate compound are incubated in prominin-1 protein-presenting wells and the amount of complex trapped in the well can be quantitated. Methods for detecting such complexes, in addition to those described above for the GST-immobilized complexes, include immunodetection of complexes using antibodies reactive with the prominin-1 protein target molecule, or which are reactive with prominin-1 protein and compete with the target molecule, as well as prominin-1-linked assays which rely on detecting an enzymatic activity associated with the target molecule.

[00358] Agents that modulate prominin-1 of the present invention can be identified using one or more of the above assays, alone or in combination. It is generally preferable to use a cell-based or cell free system first and then confirm activity in an animal or other model system. Such model systems are well known in the art and can readily be employed in this context.

[00359] In yet another aspect of the invention, prominin-1 protein can be used as a "bait protein" in a two-hybrid assay or three-hybrid assay (see, e.g., U.S. Patent No. 5,283,317; Zervos et al. (1993) Cell 72:223-232; Madura et al. (1993) J. Biol. Chem. 268:12046-12054; Bartel et al. (1993) Biotechniques 14:920-924; Iwabuchi et al. (1993) Oncogene 8:1693-1696; and Brent WO94/10300), to identify other proteins, which bind to or interact with prominin-1 and are involved in prominin-1 activity. Such prominin-1-binding proteins are also likely to be involved in the propagation of signals by prominin-1 protein or prominin-1 targets as, for example, downstream elements of a prominin-1-mediated signaling pathway. Alternatively, such prominin-1-binding proteins are likely to be prominin-1 inhibitors.

[00360] The two-hybrid system is based on the modular nature of most transcription factors, which consist of separable DNA-binding and activation domains. Briefly, the assay utilizes two different DNA constructs. In one construct, the gene that codes for prominin-1 protein is fused to a gene encoding the DNA binding domain of a known transcription factor (e.g., GAL-4). In the other construct, a DNA sequence, from a library of DNA sequences that encode an unidentified protein ("prey" or "sample") is fused to a gene that codes for the activation domain of the known transcription factor. If the "bait" and the "prey" proteins are able to interact, in vivo, forming a prominin-1-dependent complex, the DNA-binding and activation domains of the transcription factor are brought into close proximity. This proximity allows transcription of a reporter gene (e.g., LacZ) which is operably linked to a transcriptional regulatory site responsive to the transcription factor. Expression of the reporter gene can be detected and cell colonies containing the functional transcription factor can be isolated and used to obtain the cloned gene which encodes the protein which interacts with prominin-1 protein.

Array:

[00361] "Array" refers to an ordered arrangement of at least two transcripts, proteins or peptides, or antibodies on a substrate. At least one of the transcripts, proteins, or antibodies represents a control or standard, and the other transcript, protein, or antibody is of diagnostic or therapeutic interest. The arrangement of at least two and up to about 40,000 transcripts, proteins, or antibodies on the substrate assures that the size and signal intensity of each labeled complex, formed between each transcript and at least one nucleic acid, each protein and at least one ligand or antibody, or each antibody and at least one protein to which the antibody specifically binds, is individually distinguishable.

[00362] An "expression profile" is a representation of gene expression in a sample. A nucleic acid expression profile is produced using sequencing, hybridization, or amplification technologies using transcripts from a sample. A protein expression profile, although time delayed, mirrors the nucleic acid expression profile and is produced using gel electrophoresis, mass spectrometry, or an array and labeling moieties or antibodies which specifically bind the protein. The nucleic acids, proteins, or antibodies specifically

binding the protein may be used in solution or attached to a substrate, and their detection is based on methods well known in the art.

[00363] A substrate includes but is not limited to, paper, nylon or other type of membrane, filter, chip, glass slide, or any other suitable solid support.

[00364] The present invention also provides an antibody array. Antibody arrays have allowed the development of techniques for high-throughput screening of recombinant antibodies. Such methods use robots to pick and grid bacteria containing antibody genes, and a filter-based ELISA to screen and identify clones that express antibody fragments. Because liquid handling is eliminated and the clones are arrayed from master stocks, the same antibodies can be spotted multiple times and screened against multiple antigens simultaneously. For more information, see de Wildt et al. (2000) Nat. Biotechnol. 18:989-94.

[00365] The array is prepared and used according to the methods described in U.S. Patent No. 5,837,832, Chee et al., PCT application W095/11995 (Chee et al.), Lockhart, D. J. et al. (1996; Nat. Biotech. 14: 1675-1680) and Schena, M. et al. (1996; Proc. Natl. Acad. Sci. 93: 10614-10619), US Patent No. 5,807,522, Brown et al., all of which are incorporated herein in their entirety by reference.

[00366] In one embodiment, a nucleic acid array or a microarray, preferably composed of a large number of unique, single-stranded nucleic acid sequences, usually either synthetic antisense oligonucleotides or fragments of cDNAs, fixed to a solid support. The oligonucleotides are preferably about 6-60 nucleotides in length, more preferably 15-30 nucleotides in length, and most preferably about 20-25 nucleotides in length.

[00367] In order to produce oligonucleotides to a known sequence for an array, the gene(s) of interest (or an ORF identified from the contigs of the present invention) is typically examined using a computer algorithm which starts at the 5' or at the 3' end of the nucleotide sequence. Typical algorithms will then identify oligomers of defined length that are unique to the gene, have a GC content within a range suitable for hybridization, and lack predicted secondary structure that may interfere with hybridization. In certain

situations it may be appropriate to use pairs of oligonucleotides on an array. The "pairs" will be identical, except for one nucleotide that preferably is located in the center of the sequence. The second oligonucleotide in the pair (mismatched by one) serves as a control. The number of oligonucleotide pairs may range from two to one million. The oligomers are synthesized at designated areas on a substrate using a light-directed chemical process, wherein the substrate may be paper, nylon or other type of membrane, filter, chip, glass slide or any other suitable solid support as described above.

[00368] In another aspect, an oligonucleotide may be synthesized on the surface of the substrate by using a chemical coupling procedure and an ink jet application apparatus, as described in PCT application W095/251116 (Baldeschweiler et al.) which is incorporated herein in its entirety by reference.

[00369] A gene expression profile comprises the expression of a plurality of transcripts as measured by after hybridization with a sample. The transcripts of the invention may be used as elements on an array to produce a gene expression profile. In one embodiment, the array is used to diagnose or monitor the progression of disease. Researchers can assess and catalog the differences in gene expression between healthy and diseased tissues or cells.

[00370] For example, the transcript or probe may be labeled by standard methods and added to a biological sample from a patient under conditions for the formation of hybridization complexes. After an incubation period, the sample is washed and the amount of label (or signal) associated with hybridization complexes, is quantified and compared with a standard value. If complex formation in the patient sample is significantly altered (higher or lower) in comparison to either a normal or disease standard, then differential expression indicates the presence of a disorder.

[00371] In order to provide standards for establishing differential expression, normal and disease expression profiles are established. This is accomplished by combining a sample taken from normal subjects, either animal or human or nonmammal, with a transcript under conditions for hybridization to occur. Standard hybridization complexes may be quantified by comparing the values obtained using normal subjects with values from an experiment in which a known amount of a purified sequence is used. Standard values obtained in this

manner may be compared with values obtained from samples from patients who were diagnosed with a particular condition, disease, or disorder. Deviation from standard values toward those associated with a particular disorder is used to diagnose that disorder.

[00372] By analyzing changes in patterns of gene expression, disease can be diagnosed at earlier stages before the patient is symptomatic. The invention can be used to formulate a prognosis and to design a treatment regimen. The invention can also be used to monitor the efficacy of treatment. For treatments with known side effects, the array is employed to improve the treatment regimen. A dosage is established that causes a change in genetic expression patterns indicative of successful treatment. Expression patterns associated with the onset of undesirable side effects are avoided.

[00373] In another embodiment, animal models which mimic a human disease can be used to characterize expression profiles associated with a particular condition, disease, or disorder; or treatment of the condition, disease, or disorder. Novel treatment regimens may be tested in these animal models using arrays to establish and then follow expression profiles over time. In addition, arrays may be used with cell cultures or tissues removed from animal models to rapidly screen large numbers of candidate drug molecules, looking for ones that produce an expression profile similar to those of known therapeutic drugs, with the expectation that molecules with the same expression profile will likely have similar therapeutic effects. Thus, the invention provides the means to rapidly determine the molecular mode of action of a drug.

[00374] Such assays may also be used to evaluate the efficacy of a particular therapeutic treatment regimen in animal studies or in clinical trials or to monitor the treatment of an individual patient. Once the presence of a condition is established and a treatment protocol is initiated, diagnostic assays may be repeated on a regular basis to determine if the level of expression in the patient begins to approximate that which is observed in a normal subject. The results obtained from successive assays may be used to show the efficacy of treatment over a period ranging from several days to years.

EXAMPLES:

[00375] The invention is further described in the following examples, which are not intended to limit the scope of the invention. Cell lines used were obtained from the American Type Culture Collection (ATCC, Manassas, VA). The following colorectal cell lines used in this study were obtained from ATCC: Caco-2, Colo201, Colo320DM, DLD-1, HCT116, HCT15, HT29, LS123, RKO, SW480, SW620, and SW1417. Cell lines were cultured at 37°C with 5% CO_2 with appropriate media.

1. Prominin-1 mRNA Expression in Cancer Cell Lines

[00376] Expression of prominin-1 mRNA is quantitated by RT-PCR using TaqMan® technology. The Taqman® system couples a 5' fluorogenic nuclease assay with PCR for real-time quantitation. A probe is used to monitor the formation of the amplification product.

[00377] Total RNA is isolated from disease model cell lines using an RNEasy kit® (Qiagen) per manufacturer's instructions and included DNase treatment. Normal human tissue RNAs are acquired from commercial vendors (Ambion, Austin, TX; Stratagene, La Jolla, CA; BioChain Institute, Newington, NH) as were RNAs from matched disease/normal tissues.

[00378] Target transcript sequences are identified for the differentially expressed peptides by searching the BlastP database. TaqMan® assays (PCR primer/probe sets) specific for those transcripts are identified by searching the Celera Discovery System™ (CDS) database. The assays are designed to span exon-exon borders and do not amplify genomic DNA.

[00379] The TaqMan® primers and probe sequences are as designed by Applied Biosystems (AB) as part of the Assays on Demand™ product line or by custom design through the AB Assays by Design℠ service.

[00380] RT-PCR is accomplished using AmpliTaq Gold® and MultiScribe™ reverse transcriptase in the One Step RT-PCR Master Mix reagent kit (AB) according to the manufacturer's instructions. Probe and primer concentrations are 250 nM and 900 nM, respectively, in a 15 μl reaction. For each experiment, a master mix of the above

components is made and aliquoted into each optical reaction well. Eight nanograms of total RNA is the template. Each sample is assayed in triplicate. Quantitative RT-PCR is performed using the ABI Prism® 7900HT Sequence Detection System (SDS). Cycling parameters follow: 48°C for 30 min. for one cycle; 95 °C for 10 min for one cycle; and 95 °C for 15 sec, 60 °C for 1 min. for 40 cycles.

[00381] The SDS software calculates the threshold cycle (C_T) for each reaction, and C_T values are used to quantitate the relative amount of starting template in the reaction. The C_T values for each set of three reactions are averaged for all subsequent calculations

[00382] Data are analyzed to determine estimated copy number per cell. Gene expression is quantitated relative to 18S rRNA expression and copy number is estimated assuming 5 x 10^6 copies of 18S rRNA per cell (See Livak, K.J. and Schmittgen, T.D., 2001, Methods 25: 402-408; User bulletin #2: ABI Prism 7700 Sequence Detection System).

[00383] Results of prominin-1 mRNA expression analysis in cell lines are shown in Figures 1-2. As shown in Figure 1, prominin-1 mRNA is expressed at high levels in various cancer cell lines, particularly colon cancer (e.g., Caco-2, Colo201, HCT116, and HT29 cell lines) and breast cancer (e.g., MDA-MB-468) cell lines. Figure 2 shows that prominin-1 is expressed at high levels in additional cancer cell lines, particularly liver cancer cell lines (e.g., Hep 3B2).

2. Detection of Prominin-1 in Colorectal Samples by Liquid Chromatography and Mass Spectrometry (LC/MS)

[00384] Colorectal tissue samples (normal, tumor and metastatic lesions) were obtained from multiple clinical sites. Procurement of all samples was performed in an anonymised manner in strict compliance with Federally mandated ethical and legal guidelines (HIPAA) and in accordance with clinical institution ethical review board as well as the internal institutional review board. The samples for analysis were single cell suspensions prepared from surgically resected neoplastic lesions and normal adjacent tissue specimens through a series of mechanical disaggregation and enzymatic digestion steps.

[00385] Single cell suspensions were prepared from each resected sample as follows: specimens were washed in DTT for 15 min, digested with Dispase (30-60 min), then filtered twice (380 µm / 74 µm) before red blood cells were removed through addition of ACK lysis buffer. Epithelial (EpCAM), leukocyte (CD45) content and cellular viability (PI exclusion) were determined through flow cytometry analysis (LSR I, BD Biosciences, San Jose, CA). The epithelial content of both tumor and normal specimens was enriched through depletion of immune CD45 positive cells by flow cytometry or purification of Epithelial Cell Surface Antigen (ECSA/EpCam) positive cells by bead capture. Bead capture was performed using a Dynal CELLection Epithelial Enrich kit (Invitrogen, Carlsbad, CA).

[00386] For LC-MS analysis, proteins were reduced in 2.5 mM DTT for 1 hour at 37°C, and alkylated with ICAT™ reagent according to the procedures recommended by manufacturer (Applied Biosystems, Framingham, MA). The reaction was quenched by adding excess DTT. Proteins were digested using sequencing grade modified trypsin over night at 37°C followed by desalting using 3 cc Oasis HLB solid phase extraction columns (Waters, Milford, MA) and vacuum drying. Cysteine-containing peptides were purified by avidin column (Applied Biosystems, Framingham, MA). The peptides were reconstituted in buffer A (0.1% formic acid in water) and separated over a C18 monomeric column (150 mm, 150 µm i.d., Grace Vydac 238EV5, 5 µm) at a flow rate of 1.5 µl/min with a trap column. Peptides were eluted from the column using a gradient, 3%-30% buffer B (0.1% formic acid in 90% acetonitrile) in 215 min, 30%-90% buffer B in 30 min. Eluted peptides were analyzed using an online QSTAR XL system (MDS/Sciex, Toronto, ON). Peptide ion peaks from the map were automatically detected with RESPECT™ (PPL Inc., UK).

[00387] The sequence-composition of peptides detected at ≥ 4-fold higher levels in the tumor samples relative to the adjacent normal tissue, was resolved through tandem mass spectrometry and database analysis. For data analysis, peptide ion peaks of LC/MS maps from normal and tumor samples were aligned based on mass to charge ratio (m/z), retention time (Rt) and charge state (z). The list of aligned peptide ions was loaded into Spotfire™ (Spotfire Inc. Somerville, MA). Intensities were normalized before further differential analysis between tumor and normal samples. Differentially expressed ions were manually

verified before LC-MS/MS based peptide sequencing and database search for protein/protein identification.

[00388] For intensity normalization and expression analysis, a heat map was constructed by sorting the rows by the ratio of the mean intensity in the tumor samples to the mean intensity of the normal samples. Rows were only included if there was at least one MS/MS identification of an ion in the row. The display colors were determined for each row separately by assigning black to the median intensity in the row, green to the lowest intensity in the row, and red to the highest intensity (data not shown).

[00389] Using this mass spectrometry discovery procedure, a comprehensive analysis of differentially expressed cell surface proteins was performed on colorectal cancer tumor cells. The analysis involved 67 tissue samples, including tumor specimens that spanned disease stages I through IV. The majority of the normal tissue specimens included in the analysis were normal adjacent colon tissue collected during tumor resection. However, normal colon tissue from non-cancer patients was also included in the study to reduce the contribution from pre-neoplastic changes that may exist in normal adjacent tissue. 1341 peptide ions, representing some 453 distinct proteins were identified (data not shown). Several well-characterized cell-surface proteins including CarcinoEmbryonic Antigen (CEA) and EpCAM showed significantly elevated levels of expression in the population of colorectal tumors.

[00390] Prominin-1 (CD133) was identified as being dramatically over-expressed in multiple colon tumor samples. Six distinct CD133-derived peptides showed overlapping patterns of expression across the panel of resected tissue samples (data not shown). Over-expression of at least one CD133 peptide was observed in 71% of the tumor specimens analyzed and crossed stages I-IV of disease. These data indicate that CD133 is highly expressed on colorectal cancer tumor cells.

3. FACS Analysis of Primary Colorectal Tumor Samples

[00391] CD133 expression levels on primary tissue samples were quantified using the Quantum Simply Cellular System (Bangs Laboratories, Fishers, IN) and PE-conjugated

AC133 antibody (Miltenyi Biotech, Auburn, CA). Normal adjacent and cancerous colon tissues were processed into single cell suspensions, as described above. The single cell suspensions were prepared as described for proteomic analysis and stained for the epithelial marker EpCam as well as PE-conjugated AC133 antibody. Cells were analyzed by flow cytometry and the percent viable epithelial cells positive for CD133 expression were measured. Standard curve and samples were analyzed on a LSR I (BDBiosciences, San Jose CA) flow cytometer. AC133 Antibody Binding Capacity for each lineage population was calculated using geometric means and linear regression.

[00392] Greater than 50 normal-tumor tissue pairs were analyzed spanning stage I through stage IV disease (Figure 6A). The percent EpCAM+/CD133+ cells in tumors varied significantly, ranging from 0 to 58% while the mean EpCAM+/CD133+ cells in normal colon was 4.7%. Approximately 74% of the tumor tissues had increased CD133 positive cells as compared to normal adjacent colon tissue. The number of tumor cells positive for CD133 expression by flow cytometry was maintained with disease progression.

[00393] Quantitative flow cytometry was also used to compare the level of CD133 expression on colorectal cancer tumor cells to the level found on normal colon and hematopoietic cells. Frozen normal peripheral blood and bone marrow mononuclear cells were thawed using standard practices. Hematopoietic cells from both bone marrow and PBMC were stained for specific markers (CD3, CD4, CD8, CD19, CD56, CD14 and CD34; BDBiosciences) to identify CD4+ and CD8+ T-cell, NK cell, B-cell, monocyte and CD34+ stem cell populations. Simultaneously, the cells were analyzed for CD133 expression levels and compared to expression levels on normal colon and colorectal cancer tumor cells. Cells were stained with the PE-conjugated AC133 antibody and epitope copy number was determined. Standard curve and samples were analyzed on a LSR I (BDBiosciences, San Jose CA) flow cytometer. AC133 Antibody Binding Capacity for each lineage population was calculated using geometric means and linear regression.

[00394] Six out of eight tumors analyzed expressed elevated levels of CD133 compared to normal adjacent colon tissue (Figure 6B). On tumors that were positive for CD133 over-expression, the CD133 epitopes ranged from ~30,000-180,000 copies per cell while the

copy number in hematopoietic CD34+ cells from bone marrow and PBMC was only ~4,500 copies per cell. CD133 expression was not detected on the mature hematopoietic cell types tested. These data confirm over-expression of CD133 in CRC tumor cells relative to normal colon and demonstrate that CD133 expression levels are significantly higher on tumor cells than on normal colon and hematopoietic stem cells.

4. Quantitative FACS Analysis of Prominin-1 Expressing Cell Lines

[00395] Cell surface CD133 expression levels were quantified on cell lines with QIFIKIT flow cytometric indirect immunofluorescence assay (Dako A/S) using AC133 as the primary antibody. The cell lines examined were Hep3B, HepG2, Su86.86, Capan-1, Capan-2, KatoIII, AGS, HS174T, Caco-2, HCT116, HCT15, DLD1, Colo320, RKO, Colo201, HT29, LS123, LoVo and SW620. Briefly, cells were detached with versene or trypsin and washed once with complete media then PBS. 5×10^5 cells/sample were incubated with saturating concentration (10 µg/ml) of primary antibody for 60 minutes at 4°C. After washes, FITC-conjugated secondary antibody (1:50 dilution) was added for 45 minutes at 4°C. QIFIKIT standard beads were simultaneously labeled with the secondary antibody. Binding of antibodies was analyzed by flow cytometry and specific antigen density was calculated by subtracting background antibody equivalent from antibody-binding capacity based on a standard curve of log mean fluorescence intensity versus log antigen binding capacity.

[00396] Referring to Figure 7A, several cell lines that were positive for CD133 expression were identified. Notably, Hep3B, a metastatic hepatocellular cell line, had relatively high levels of CD133 (78,000 sites/cell) compared to the pancreatic cell lines (Su86.86, CAPAN-1 and CAPAN-2) and another hepatocellular line (HepG2).

5. Expression Validation by Immunohistochemistry (IHC) in Tissue Sections

Tissue Sections

[00397] Paraffin embedded, fixed tissue sections are obtained from a panel of normal tissues (adrenal, bladder, lymphocytes, bone marrow, breast, cerebellum, cerebral cortex, colon, endothelium, eye, fallopian tube, small intestine, heart, kidney [glomerulus, tubule],

liver, lung, testes and thyroid) as well as 30 tumor samples with matched normal adjacent tissues from pancreas, lung, colon, prostate, ovarian and breast. In addition, other tissues are selected for testing such as bladder, renal, hepatocellular, pharyngeal and gastric tumor tissues. Replicate sections are also obtained from numerous tumor types (bladder cancer, lung cancer, breast cancer, melanoma, colon cancer, non-hodgkins lymphoma, endometrial cancer, ovarian cancer, head and neck cancer, prostate cancer, leukemia [ALL and CML] and rectal cancer). Sections are stained with hemotoxylin and eosin and histologically examined to ensure adequate representation of cell types in each tissue section.

[00398] An identical set of tissues is obtained from frozen sections and is used in those instances where it is not possible to generate antibodies that are suitable for fixed sections. Frozen tissues do not require an antigen retrieval step.

Paraffin Fixed Tissue Sections

[00399] Hemotoxylin and eosin staining of paraffin embedded, fixed tissue sections. Sections are deparaffinized in three changes of xylene or xylene substitute for 2-5 minutes each. Sections are rinsed in two changes of absolute alcohol for 1-2 minutes each, in 95% alcohol for 1 minute, followed by 80% alcohol for 1 minute. Slides are washed well in running water and stained in Gill solution 3 hemotoxylin for 3-5 minutes. Following a vigorous wash in running water for 1 minute, sections are stained in Scott's solution for 2 minutes. Sections are washed for 1 minute in running water and then counterstained in eosin solution for 2-3 minutes, depending upon the desired staining intensity. Following a brief wash in 95% alcohol, sections are dehydrated in three changes of absolute alcohol for 1 minute each and three changes of xylene or xylene substitute for 1-2 minutes each. Slides are coverslipped and stored for analysis.

Optimization of Antibody Staining

[00400] For each antibody, a positive and negative control sample are generated using data from the ICAT analysis of the cancer cell lines/tissues. Cells are selected that are known to express low levels of a particular target as determined from the ICAT data. This cell line is the reference normal control. Similarly, a cancer cell line that is determined to over-express the target is selected.

Antigen Retrieval

[00401] Sections are deparaffinized and rehydrated by washing 3 times for 5 minutes in xylene, two times for 5 minutes in 100% ethanol, two times for 5 minutes in 95% ethanol, and once for 5 minutes in 80% ethanol. Sections are then placed in endogenous blocking solution (methanol + 2% hydrogen peroxide) and incubated for 20 minutes at room temperature. Sections are rinsed twice for 5 minutes each in deionized water and twice for 5 minutes in phosphate buffered saline (PBS), pH 7.4.

[00402] Alternatively, where necessary, sections are deparrafinized by High Energy Antigen Retrieval as follows: sections are washed three times for 5 minutes in xylene, two times for 5 minutes in 100% ethanol, two times for 5 minutes in 95% ethanol, and once for 5 minutes in 80% ethanol. Sections are placed in a Coplin jar with dilute antigen retrieval solution (10 mM citrate acid, pH 6). The Coplin jar containing slides is placed in a vessel filled with water and microwaved on high for 2-3 minutes (700 watt oven). Following cooling for 2-3 minutes, steps 3 and 4 are repeated four times (depending on tissue), followed by cooling for 20 minutes at room temperature. Sections are then rinsed in deionized water (two times for 5 minutes), placed in modified endogenous oxidation blocking solution (PBS + 2% hydrogen peroxide), and rinsed for 5 minutes in PBS.

Blocking and Staining

[00403] Sections are blocked with PBS/1% bovine serum albumin (PBA) for 1 hour at room temperature followed by incubation in normal serum diluted in PBA (2%) for 30 minutes at room temperature to reduce non-specific binding of antibody. Incubations are performed in a sealed humidity chamber to prevent air-drying of the tissue sections. (The choice of blocking serum is the same as the species of the biotinylated secondary antibody.) Excess antibody is gently removed by shaking and sections covered with primary antibody diluted in PBA and incubated either at room temperature for 1 hour or overnight at 4°C. (Care is taken that the sections do not touch during incubation). Sections are rinsed twice for 5 minutes in PBS, shaking gently. Excess PBS is removed by gently shaking. The sections are covered with diluted biotinylated secondary antibody in PBA and incubated for 30 minutes to 1 hour at room temperature in the humidity chamber. If using a monoclonal primary antibody, addition of 2% rat serum is used to decrease the background on rat tissue

sections. Following incubation, sections are rinsed twice for 5 minutes in PBS, shaking gently. Excess PBS is removed and sections incubated for 1 hour at room temperature in Vectastain ABC reagent (as per kit instructions). The lid of the humidity chamber is secured during all incubations to ensure a moist environment. Sections are rinsed twice for 5 minutes in PBS, shaking gently.

Develop and Counterstain

[00404] Sections are incubated for 2 minutes in peroxidase substrate solution that is made up immediately prior to use as follows:

- 10 mg diaminobenzidine (DAB) dissolved in 10 ml of 50 mM sodium phosphate buffer, pH 7.4.
- 12.5 microliters 3% $CoCl_2/NiCl_2$ in deionized water
- 1.25 microliters hydrogen peroxide

[00405] Slides are rinsed well three times for 10 min in deionized water and counterstained with 0.01% Light Green acidified with 0.01% acetic acid for 1-2 minutes depending on the desired intensity of counterstain.

[00406] Slides are rinsed three times for 5 minutes with deionized water and dehydrated two times for 2 minutes in 95% ethanol; two times for 2 minutes in 100% ethanol; and two times for 2 minutes in xylene. Stained slides are mounted for visualization by microscopy.

[00407] As shown in Figure 3, IHC demonstrated overexpression of prominin-1 by two pathology grades in multiple tumor specimens, including those from colon (over-expressed in 60% of tumors), liver (50%), non-small cell lung carcinoma (40%), lung squamous (30%), melanoma (30%), ovary (40%), pancreas (67%), pharynx (60%), kidney (10%), and prostate (40%) tumors. IHC also demonstrated overexpression of prominin-1 by one pathology grade in bladder tumor (over-expressed in 50% of tumors) and breast tumor (30%) specimens.

<u>6. Expression Validation of Colorectal Cancer Specimens by IHC in Tissue Sections</u>

[00408] In a variation of the IHC expression validation described above, 29 formalin fixed paraffin embedded (FFPE) tissues were deparaffinized and processed for antigen retrieval using EZ-retriever system (BioGenex, San Ramon, CA). 29 primary colorectal tumor cases with varying degrees of differentiation were analyzed, eight of which had matching metastatic tumors. The colorectal cancer specimens were received from Asterand (Detroit, MI) or Fox Chase Cancer Center (Philadelphia, PA). The immunohistochemical analysis was performed using two anti-CD133 antibodies, Ab5558 (Abcam) and AC133 (Miltenyi), as primary antibodies.

[00409] EZ-antigen Retrieval common solution was used for deparaffinization and EZ-retrieval citrate-based buffer was used for antigen retrieval. Samples were pre-blocked with non-serum protein block (DAKO A/S, Glostrup, Denmark) for 15 min. Primary antibodies were incubated overnight at room temperature. Monoclonal antibodies AC133 and control mouse IgG MOPC21 (Sigma, St. Louis, MO) were used at 5.0 μg/ml. Monoclonal antibody against CD133 (ab5558) (Abcam) was used at 2.5 μg/ml. Envision Plus system HRP (DAKO) was used for detection with diaminobenzidine (DAB) as substrate for horseradish peroxidase. Slides were then scored manually using a Zeiss Axiovert 200M microscope (Carl Zeiss Microimaging, Thornwood, NY). Colorectal cell lines expressing CD133 were used as positive controls for staining. Representative images were acquired using 40x objective (400x magnification).

[00410] Seventy-six percent (22 out of 29) of the cases demonstrated positive CD133 expression. The immunostaining intensity with both antibodies varied from weak to strong and was typically localized to the cell membranes and the luminal spaces of neoplastic gland-like structures, and rarely the extracellular spaces adjacent to neoplastic cells (mucinous type adenocarcinoma). The nature of the luminal immunostaining is unclear, but does appear to be specific for CD133. The distribution of immunostaining varied from 5 to 90% of tumor cells within a given tumor sample. There did not appear to be a clear correlation between tumor differentiation and immunostaining intensity or distribution (see Table 1). However, the moderately and well differentiated tumors appeared to show a trend toward more intense and greater distribution of immunostaining. Ab5558 generally exhibited very slightly more intense immunostaining and percentage distribution, and

demonstrated cytoplasmic immunostaining not observed with tumors immunostained using AC133. These data confirm expression of CD133 in colorectal cancer CRC tumor cells and suggest that expression levels do not correlate with the level of tumor differentiation. .

[00411] In metastasis-matched colon tumor samples, most (7 of 8) cases revealed weak to strong immunostaining for CD133, and exhibited the typical membranous and luminal localization similar to the primary colorectal cases. The percent distribution of primary tumor cells was relatively high in most cases (50-100%). There was good correlation between primary tumor and associated metastasis in terms of immunostaining intensity and percent distribution. Commonly, the immunostaining of metastatic sites was of equal or greater intensity than the associated primary tumor. Also, in most cases the level of differentiation between primary and metastasis correlated well. There was no clear correlation between tumor differentiation and immunostaining intensity and percent distribution.

7. Identification of Prominin-1 Expression by IHC in Solid Tumor Samples

[00412] To determine the expression of prominin-1 (CD133) in various solid tumors, two monoclonal antibodies, AC133.1 (Miltenyi) and Ab5558, (Abcam, Cambridge, MA), were used for detection of CD133 in formalin-fixed paraffin-embedded (FFPE) samples by immunohistochemistry. For the initial survey, a panel of common cancer tissue microarray (TMA) (12 samples per tumor type) was used for the analysis. The panel included samples of lung, breast, ovary, colon, melanoma, pancreatic,. kidney, head and neck, liver, and prostatic carcinomas.

[00413] The formalin-fixed paraffin-embedded (FFPE) tissue microarrays (TMAs) were obtained from commercial sources (TriStar, Rockville, MD; USBiomax, Rockville, MD; Imgenex, San Diego, CA; Petagen/Abxis, Seoul, Korea). Slides were deparaffinized and processed for antigen retrieval using EZ-retriever system (BioGenex, San Ramon, CA). EZ-antigen Retrieval common solution was used for deparaffinization and EZ-retrieval citrate-based buffer was used for antigen retrieval. Samples were pre-blocked with non-serum protein block (Dako A/S, Glostrup, Denmark) for 15 minutes. Primary antibodies were incubated overnight at room temperature. MAb CD133/1 (Miltenyi, Auburn, CA)

and control MAb IgG were used at 5.0 µg/ml, whereas anti-CD133 MAb, Ab5558, (Abcam, Cambridge, MA) was used at 2.5 µg/ml. Envision Plus system HRP (Dako A/S) was used for detection with diaminobenzidine (DAB) as substrate for horseradish peroxidase. Slides were then scored manually using a Zeiss Axiovert 200M microscope (Carl Zeiss Microimaging, Thornwood, NY). A Caco-2 colorectal cell line expressing CD133 was used as positive control for staining. Images were acquired using a 40x objective (400x magnification).

[00414] Several tumor types showed weak to strong staining of prominin-1 expression, including pancreatic (4/12 samples) and hepatocellular carcinomas (12/12 samples). Additional samples for each tumor type were studied using tissue microarrays specific for pancreatic, gastric, renal, prostatic and hepatocellular carcinomas. The intensity of the staining for CD133 ranged from weak (1-2+) to strong (3-4+). A good concordance was observed between the reactivity of the two antibodies to the samples, with a difference in intensity of the staining in some cases (data not shown), although Ab5558 gave more consistent IHC staining than AC133.1. The luminal pattern of expression is also highly characteristic of the CD133 staining pattern observed in other solid tumors such as colorectal cancers.

[00415] The results of the analysis are summarized in Table 2. A high percentage (≥50%) of CD133+ tumor cases was observed in gastric (55%), pancreatic (58%), and cholangiocarcinomas (biliary type of liver cancer) (67%) samples. CD133 expression was also detected in 29% of renal cell carcinoma cases. There were some metastatic tumors (20-30 cases primarily of gastric and colorectal origin) included in the gastric and liver TMAs which also showed a high percentage (≥50%) of CD133+ cases. In normal tissues, CD133 mild to strong apical membrane staining of pancreatic acinar and ductal epithelium, biliary ducts of liver and tubular epithelium of kidney was observed (data not shown).

Table 2

Tumor Type	Number of cases	Percentage CD133+

Primary		
Gastric adenocarcinomas	60	55
Pancreatic ductal adenocarcinomas	31	58
Renal cell carcinomas	31	29
Cholangiocarcinomas (bile duct)	12	67
Prostatic adenocarcinomas	39	13
Metastatic		
Liver (predominantly colonic in origin)	30	63
Gastric	20	50

8. Preparation of Antibody-Drug Conjugates

[00416] The hybridoma cell line (AC133.1, ATCC, Manassas, VA), producing the murine anti-CD133 antibody (AC133), was grown as recommended and media collected for antibody purification. Antibodies were purified using MabSelect Protein A column (Amersham, Piscataway, NJ). MAb AC133 in 50 mM sodium borate, 50 mM NaCl, and 1 mM DTPA (Diethylenetriaminepentaacetic acid) pH 8.0 was partially reduced with 2.5 equivalents of tris(2-carboxyethyl)phosphine hydrochloride at 37°C for 1 h to yield about 5.3 thiols per antibody. The mixture was cooled to 0°C and partially reoxidized with 0.48 equivalents of 5,5'-dithiobis-(2-nitrobenzoic acid) (DTNB) to about 4.4 thiols per antibody. This mixture was divided in three equal portions of and alkylated with 1.5 equivalents per thiol with one of the following auristatin drug-linkers for 30 min: maleimidocaproyl-valine-citrulline-p-aminobenzoyl-MMAE (vcMMAE), maleimidocaproyl-valine-citrulline-p-aminobenzoyl-MMAF (vcMMAF) and maleimidocaproyl-MMAF (mcMMAF) (PCT Publication No. WO 2005/081711. Excess drug-linker was quenched with excess N-acetyl-cysteine and the entire mixture purified on a centrifugal–S-Fast Flow cation exchange cartridge in 30 mM sodium acetate (pH 5.0) and eluted with 3 x PBS. The eluted conjugates were diluted with water to 1 x PBS, concentrated, and filter-sterilized (0.2 μm).

[00417] For conjugation to a fluorophore, AC133 or AC133-drug conjugates in 50 mM sodium borate, 50 mM NaCl pH 8.0 were reacted with 6 equivalents of Alexa Fluor 594 N-hydroxysuccinimide ester (Invitrogen, Carlsbad, CA) at 25°C for 1 h. The mixture was purified on a PD-10 column equilibrated with PBS. A similar procedure was used for

conjugating AC133 or AC133-drug conjugates to Alexa Fluor 488 N-hydroxysuccinimide ester.

9. *In vitro* Assays in Cell Lines Using RNAi

RNAi and RNAi Transfections

[00418] Pooled synthetic small interfering RNA (SMARTpool siRNA) specific for CHK1 (MU-003255) and scrambled negative control (D-001216-13) were purchased from Dharmacon (Lafayette, CO). CD133 expressions were knocked down by transfection with CD133 siRNA (Dharmacon, USA) with duplex 1 directed against the sequence gccagaaactgtaatctta (nucleotides 313-331 of SEQ ID NO:11, see Figure 5), duplex 2 directed against the sequence taacaaatgtggtggagaa (nucleotides 445-463 of SEQ ID NO:11, see Figure 5) and duplex 3 directed against the sequence gctcagaacttcatcacaa (nucleotides 2204-2222 of SEQ ID NO:11, see Figure 5).

[00419] For siRNA transfection, HT29 colon adenocarcinoma cells were seeded into 96 well tissue culture plates at a density of 2,500 cells per well 24 hours before transfection. Culture medium was removed and 50µl of reaction mix containing siRNA (final concentration 1 to 100 nM) and 0.4 µl of DharmaFECT4 (Dharmacon, Lafayette, CO) diluted in Opti-MEM was added to each well. An equal volume of complete medium followed and the cells were then incubated at 5% CO_2 at 37 °C for 1 to 4 days.

mRNA and Protein Knockdowns

[00420] Knockdown of CD133 mRNA levels was monitored by Q-PCR 1 day after siRNA transfection by using a TaqMan® Gene Expression Assay for PROM1 (Hs01009257_m1, Applied Biosystems, ABI, Foster City, CA). RT-PCR was accomplished in a one-step reaction by using M-MLV reverse transcriptase (Promega, Madison, WI) and AmpliTaq Gold® (ABI) and analyzed on the ABI Prism® 7900HT Sequence Detection System (ABI). Relative gene expression was quantitated by the ΔΔCt method (User Bulletin #2, ABI.) with 18S rRNA serving as the endogenous control.

[00421] Protein knockdown was monitored by FACS 4 days after transfection by using anti-CD133 antibody from Miltenyi Biotech (clone AC133, Auburn, CA). The samples were ran on a LSR flow cytometer (BD Biosciences, San Jose, CA) and live cells

were monitored by using PI exclusion (50 µg/ml PI, 2.5 Units/ml Rnase A, 0.1% Triton X-100 in D-PBS). The data was analyzed by using CellQuest software.

Cell Proliferation - Alamar Blue

[00422] Cell growth was assessed 4 days after transfection by adding a 1:10 dilution of alamar blue reagent (Invitrogen, Carlsbad, CA) and incubated for 2 h at 37 °C. Analysis was performed on a Spectrafluor Plus (Tecan, Durham, NC) set at excitation wavelength of 530 nm and emission wavelength of 595 nm.

[00423] Knockdown of CD133 mRNA inhibited proliferation of colon cancer cells. Q-PCR indicates that HT29 colon adenocarcinoma cells express greater than 100 copies of CD133 mRNA per cell (results not shown). Following the transfection of HT29 cells with three individual siRNA duplexes directed against CD133 at 100 nM, all three siRNA duplexes decreased both CD133 mRNA and protein levels (Figures 4A-B). For each duplex, the level of protein knockdown corresponded to the level of knockdown at the mRNA level. The three siRNA duplexes were then titrated down to 1 nM and effects on cell proliferation were monitored. A dose-dependent inhibition of proliferation was observed with all three duplexes (Figure 4C). This is in contrast to the scrambled negative control siRNA, which did not inhibit cell growth. These results suggest a functional role for CD133 in the proliferation of colon cancer cells.

10. *In vitro* Assays in Cell Lines Using Antibodies

Cytotoxicity Assays

[00424] Cytotoxicity was measured using a Resazurin (Sigma, MO) dye reduction assay, as described previously (McMillian, M.K. et al., 2002, Cell Biol. Toxicol. 18:157-173). Briefly, cells were plated at 1,000-5,500 cells/well in 96 well plates, allowed to attach to the plates for 18 hours before addition of fresh media with or without antibody-drug conjugates or antibody. After 96 - 144 hours of exposure to antibody or antibody-drug conjugates, resazurin was added to cells to a final concentration of 50 µM. Cells were incubated for 2-6 hours depending on dye conversion of cell lines, and dye reduction was measured on a Fusion HT fluorescent plate reader (Packard Instruments, Meridien, CT) with excitation and emission wavelengths of 530 nm and 590 nm, respectively. The IC_{50}

value is defined here as the drug concentration that results in 50% reduction in growth or viability as compared with untreated control cultures.

Proliferation Assays

[00425] To measure cell proliferation, cells were plated, grown and treated as for cytotoxicity assay (above) in 96 well plates. After 96-144 h of treatment, 0.5 µCi/well ^{3}H-Thymidine (PerkinElmer, 6.7 Ci/mmol) was added to cells and incubated for 4-6 h at 37 °C, 5% CO_2 in an incubator. To lyse cells, plates were frozen overnight at -20 °C then cell lysates were harvested using FilterMate (Packard Instrument, Meridien, CT) into 96 well filter plate. Radioactivity associated with cells was measured on TopCount (Packard) scintillation counter.

[00426] *Effect of Antibody-Drug Conjugates on Recombinant Prominin-1 Expressing Cells*

[00427] AC133 antibody drug conjugates (ADCs) to the anti-mitotic drugs monomethyl auristatin E or F (MMAE or MMAF) either with or without a cathepsin B cleavable dipeptide linker val-cit (vc) were prepared as described above. These ADCs were used for in vitro assays measuring cytotoxicity or cell proliferation, as described above.

[00428] A mammalian expression vector was used to transfect HEK293 cells with the human CD133 gene and a stable clone expressing very high levels of CD133 (~2 x 10^5 copies CD133/cell) was isolated and used for the cytotoxicity assays (Figure 8A). Human Embryonic Kidney 293 cells were grown in Dulbecco's Modified Eagle medium supplemented with 10% FBS. Cells were plated 24 h before transfection. Cells were transfected with an expression construct for CD133 using Lipofectamine 2000 (Invitrogen, Carlsbad, CA) according to the manufacturer's instructions. Stable transfectants were selected using 800 µg/ml Geneticin and screened using flow cytometry with the AC133 monoclonal antibody (Miltenyi Biotech, Auburn, CA).

[00429] Both AC133mcMMAF (no dipeptide linker) and AC133vcMMAF (with cleavable linker) were very effective in killing the CD133-transfected cells with IC_{50} values

of ~1 ng/ml. AC133vcMMAE was also cytotoxic, with an IC_{50} value of 6 ng/ml. The untransfected 293 cells were not affected by the ADCs (IC50 > 10 μg/ml).

[00430] The effects of anti-CD133 drug conjugates on viability and cell proliferation of established colorectal cell lines was also determined (see Figure 8). The viability and growth of Caco-2 cells (~2 x 10^5 CD133/cell) were effectively inhibited by AC133vcMMAF. Results from the cytotoxicity assay in Caco-2 cells showed an IC_{50} value of 3 ng/ml for AC133vcMMAF (Figure 8B). Growth inhibition was observed with AC133vcMMAF with an IC_{50} value < 1 ng/ml) while both AC133mcMMAF and AC133vcMMAE were less effective (IC50 values of 100 ng/ml and ~1 μg/ml, respectively, (Figure 8C). HCT116 cells which have low levels of CD133 (< 2 x 10^4 molecules/cell) are not sensitive to the growth inhibitory effects of any of the AC133 drug conjugates even though the control ADC demonstrated that the cells are sensitive to the drug upon internalization (Figure 8D).

Effect of Antibody-Drug Conjugates on Prominin-1 Expressing Cell Lines

[00431] MAb AC133 was conjugated to the anti-tubulin drug MMAE or MMAF (4 drugs/antibody), as described above. Surprisingly, the antibody drug conjugate was internalized. The cytotoxicity of the AC133 antibody conjugated to MMAF was demonstrated using a rezasurin dye conversion assay in Hep3B cells (see Figure 7B). The IC_{50} value for AC133-vcMMAF4 is 0.35 nM while AC133-mcMMAF4 is 6.0 nM. In contrast, the AC133 drug conjugates did not have a significant cytotoxic effect on the pancreatic cell line Su86.86, which has 35,000 CD133 sites/cell. The positive control anti-transferrin receptor antibody OKT9-vcMMAF (loaded with 8 drugs/antibody) showed cytotoxic effects on Su86.86.

[00432] When cell proliferation was measured using ^{3}H-Thymidine incorporation, the inhibitory effect of AC133-vcMMAF was observed in both Hep3B and Su86.86, with Hep3B being more sensitive to the ADC. AC133-mcMMAF and AC133-vcMMAE also inhibited cell proliferation of Hep3B cells at higher concentrations compared to AC133-vcMMAF. Therefore, in Hep3B, the AC133 drug conjugates inhibited both cell viability and proliferation while in Su86.86, AC133-vcMMAF inhibited cell proliferation alone. In

cell lines with lower CD133 expression such as CAPAN-2 (13000 sites/cell) and HCT116 (18000 sites/cell), the AC133 drug conjugates did not inhibit viability or proliferation of the cells (data not shown).

Subcellular Localization of Prominin-1 Antibody Drug Conjugates

[00433] Immunofluorescence studies were employed to track the localization of the ADC. Cells were grown in cover slip bottom chamber slides to about 75% confluence and media was changed after 48 h. Antibodies conjugated to Alexa Fluor 594 were added to the cells at 1 μg/ml and immunofluorescence was performed using a Zeiss Axiovert 200M fluorescence microscope (Carl Zeiss Microimaging, Thornwood, NY). After 1-2 days, cells were fixed and permeabilized with paraformaldehyde/saponin as provided in the Cytofix/Cytoperm kit (BD Biosciences, San Jose, CA) and then stained with anti-CD107a-FITC (lysosomal marker) (BD Biosciences). Nuclei were stained using 4′,6-diamidino-2-phenylindole (DAPI, Roche, Switzerland). Cells were mounted with ProLong Antifade solution (Invitrogen/Molecular Probes). Images were obtained using the Zeiss Axiovert 200M using 63x oil immersion objective with Apotome for optical sectioning.

[00434] Alexa-Fluor labeled ADCs demonstrate that AC133vcMMAF can be found intracellularly in Caco-2 cells (data not shown). Using a FITC-labeled lysosomal marker, co-localization of some of the ADC signal within the lysosome in Caco-2 cells was observed, where the cleavage of the dipeptide linker is most effective for the release of the drug (data not shown). In HCT116 cells, there was minimal signal observed for the ADC and no apparent colocalization with the lysosomal marker, suggesting that the lack of ADC activity on HCT116 cells was due to ineffective internalization (data not shown).

[00435] AC133-vcMMAF was observed to efficiently internalize in both Hep3B and Su86.86 cells, but the pattern of colocalization with Caveolin-1 marker differed. In Hep3B, there was less colocalization with Caveolin-1 (more of a vesicular staining pattern) while in Su86.86 most of the signal for the ADC had a good colocalization with Caveolin-1 within 24 hr. A FITC-conjugated anti-CD107a was used to determine colocalization of the ADC with the lysosomal marker at 48 hours (data not shown). Some overlap in both Hep3B and Su86.86 was observed, indicating colocalization of the AC133-vcMMAF with the

lysosomal marker at 48 hours but with much less ADC in Su86.86. Some ADC was found in the peripheral area or plasma membranes in Hep3B which is not detected in the Su86.86 cells.

11. *In Vivo* Studies Using Antibody-Drug Conjugates

[00436] The expression of CD133 in vivo in Hep3B tumor xenografts was analyzed using both FACS analysis and IHC. Hep3B cells express CD133 (see Figure 9A). SCID mice were implanted (subcutaneous injection) with Hep3B cells (1 x 10^7 cells) and established subcutaneous tumors (~100 mm^3) were treated with MAb AC133 or its drug conjugates or mouse IgG1 control (11G1) drug conjugates. Groups of mice (5 animals per group) were treated with the antibody alone or an antibody-drug conjugate, as follows: AC133 (10 mg/kg), AC133-vcMMAF4 (1.0 or 3.0 mg/kg), AC133-mcMMAF4 (3.0 or 10 mg/kg), and isotype control mouse IgG1 antibody drug conjugates (11G1-vcMMAF4 at 1.0 or 3.0 mg/kg or 11G1-mcMMAF4 at 3.0 or 10 mg/kg). The first dosing was administered intraperitoneal every 4 days for a total of 4 doses (q4d x 4).

[00437] Median tumor volume and weight of mice were monitored and tumors were collected when the tumor volume reached 1000 mm^3 for further analysis of CD133 expression by FACS or immunohistochemistry. All animal procedures were performed under a protocol approved by the Institutional Animal Care and Use Committee.

[00438] Surprisingly, the results indicate the antibody drug conjugate was internalized and effective against tumor. Within 2 days of the last dose, the Hep3B tumors in the mice treated with AC133-vcMMAF4 at 3.0 mg/kg showed a pronounced decrease in tumor size (see Figure 9B). The group treated with a lower dose of AC133-vcMMAF4 (1.0 mg/kg) showed a delayed growth of the tumor compared to the other treated groups.

[00439] A second dosing of AC133-vcMMAF (3 mg/kg) was also performed on one group of mice, as indicated in Figure 9B. After the second dosing began, one mouse was sacrificed based on tumor size.

12. Cell Transformation Assay: Soft Agar Assay for Colony Formation

[00440] The soft agar assay is the most stringent assay for detecting malignant transformation of cells using *in vitro* methods. Upon neoplastic transformation, cells have a reduced requirement for cell-to-cell contact and extracellular growth promoting factors due to accumulation of a series of genetic and epigenetic changes. The soft agar assay is commonly employed to monitor anchorage-independent growth of cells. This *in vitro* cell transformation is a reasonably good predictor of *in vivo* carcinogenesis activity.

[00441] The soft agar assay was performed according to a published report (Hudziak *et. al., Proc. Natl. Acad. Sci.*, 1987, 84, p. 7159) with some modifications. Briefly, plasmid DNAs encoding CD133 (pDEST40-CD133v2) or an empty vector (pDEST40) under a CMV promoter were introduced into murine NIH 3T3 cells by traditional methods. Murine NIH 3T3 cells are highly anchorage-dependent and exhibit low spontaneous transformation. Forty-eight hours post-transfection, the cells were passaged into selective medium containing Geneticin (G418) at 500 μg/ml and single cell clones were expanded. The efficiency of colony formation in soft agar was determined by plating 5000 cells in 1.5 ml of 0.4% top agar over 2 ml of 0.8% bottom agar in six wells of 6-well tissue culture plate. After 2-4 weeks, the colonies were stained with cell staining solution (Chemicon, CA) and counted.

[00442] No colonies were observed in wells with expression vector alone, whereas in the wells plated with cells expressing CD133, multiple clear colonies were evident (Figure 10). The experiment was performed in duplicates and was repeated twice with similar results.

[00443] All publications and patents mentioned in the above specification are herein incorporated by reference. Various modifications and variations of the described method and system of the invention will be apparent to those skilled in the art without departing from the scope and spirit of the invention. Although the invention has been described in connection with specific preferred embodiments, it should be understood that the invention as claimed should not be unduly limited to such specific embodiments. Indeed, various modifications of the above-described modes for carrying out the invention, which are

obvious to those skilled in the field of molecular biology or related fields, are intended to be within the scope of the following claims.

CLAIMS

1. A method for detecting a disease or disease recurrence in a subject, the method comprising:

determining a test level or test activity of prominin-1 in a disease sample from the subject; and

determining a control level or control activity in a corresponding sample from a healthy subject, wherein the disease is related to abnormal expression or function of prominin-1, and wherein a difference in the test level or test activity in the sample from the subject compared to the control level or control activity in the sample from the healthy subject is indicative of the presence of the disease.

2. The method of claim 1, wherein the level of the prominin-1 is determined using an antibody that specifically binds to an antigenic region of prominin-1.

3. The method of claim 1, wherein the prominin-1 comprises an amino acid sequence selected from the group consisting of SEQ ID NOS:1-5.

4. The method of claim 1, wherein the level of a nucleic acid molecule encoding prominin-1 is determined.

5. A method for monitoring treatment of a disease in a subject, wherein the disease is related to abnormal expression or function of prominin-1, the method comprising:

determining a first test level or a first test activity of prominin-1 in a sample from the subject prior to the treatment;

determining a second test level or a second test activity of prominin-1 in a sample from the subject subsequent to the treatment;

determining a control level or control activity of prominin-1 in a corresponding sample from a healthy subject; and

comparing the second test level or second test activity of prominin-1 in the sample from the subject to the control level or control activity of prominin-1, wherein a change in which the second test level or second test activity is closer to the control level or control activity than the first test level or first test activity is to the control level or control activity is indicative of successful treatment.

6. A pharmaceutical composition comprising an antagonist to prominin-1 and a pharmaceutically acceptable excipient.

7. The pharmaceutical composition of claim 6, wherein the antagonist is an anti-prominin-1 antibody.

8. The pharmaceutical composition of claim 7, wherein the antagonist is an anti-prominin-1 antibody conjugated to a therapeutic agent.

9. The pharmaceutical composition of claim 8, wherein the therapeutic agent is a cytotoxic drug, a cytostatic drug or an immunomodulatory agent.

10. The pharmaceutical composition of claim 9, wherein the therapeutic agent is MMAF or MMAE.

11. The pharmaceutical composition of claim 6, wherein the antagonist is an anti-sense nucleic acid molecule or an RNAi molecule that inhibits the translation or transcription of a nucleic acid molecule that codes for the prominin-1.

12. The pharmaceutical composition of claim 6, wherein the prominin-1 comprises an amino acid sequence selected from the group consisting of SEQ ID NOS:1-5.

13. A method for treating a disease, wherein the disease is related to abnormal expression or function of prominin-1 in a disease cell, the method comprising administering to a patient in need thereof an effective amount of the pharmaceutical composition of claim 6.

14. A method of inhibiting cell growth or proliferation comprising contacting cells with an anti-prominin-1 antibody.

15. The method of claim 14, wherein the anti-prominin-1 antibody is conjugated to a therapeutic agent.

16. The method of claim 15, wherein the therapeutic agent is a cytotoxic drug, a cytostatic drug or an immunomodulatory agent.

17. The method of claim 16, wherein the therapeutic agent is MMAF or MMAE.

18. A method of inhibiting cell growth or proliferation comprising contacting cells with the pharmaceutical composition of claim 11.

19. A method of screening for an agent that modulates the activity or expression of prominin-1 protein, the method comprising:

(i) contacting a candidate agent with a preparation selected from the group consisting of prominin-1 protein, a cell that expresses prominin-1 protein, and a cell-free preparation that expresses prominin-1 protein; and

(ii) assaying for activity or expression of prominin-1 protein, wherein a change in the activity or expression of prominin-1 protein in the presence of the agent relative to the activity or expression of prominin-1 protein in the absence of the agent indicates that the agent modulates the activity or expression of prominin-1 protein.

20. The method of claim 1, wherein the disease is selected from the group consisting of colon cancer, breast cancer, bladder cancer, kidney cancer, liver cancer, lung cancer, melanoma, ovarian cancer, pancreatic cancer, pharyngeal cancer, gastrointestinal cancer, and prostate cancer.

21. The method of claim 14, wherein the cells are cancer cells selected from the group consisting of colon cancer cells, breast cancer cells, bladder cancer cells, kidney cancer cells, liver cancer cells, lung cancer cells, melanoma cells, ovarian cancer cells,

pancreatic cancer cells, pharyngeal cancer cells, gastrointestinal cancer cells, and prostate cancer cells.

22. An antibody drug conjugate having the following formula:

$$Ab\text{-}(D)_p$$

or a pharmaceutically acceptable salt or solvate thereof;

wherein,

Ab is an Antibody Unit by binds specifically to CD133;

p ranges from 1 to about 20; and

D is a Drug Unit.

23. The antibody drug conjugate of claim 22, having the following formula:

$$Ab\text{-}(Aa\text{-}Ww\text{-}Yy\text{-}D)_p$$

or a pharmaceutically acceptable salt or solvate thereof, wherein:

Ab is an antibody,

A is a Stretcher unit,

a is 0 or 1,

each W is independently an Amino Acid unit,

w is an integer ranging from 0 to 12,

Y is a Spacer unit, and

y is 0, 1 or 2,

p ranges from 1 to about 20, and

D is a Drug Unit.

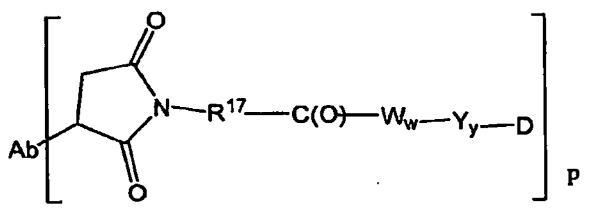
24. The antibody drug conjugate of claim 22 or 23, wherein the Drug Unit comprises a cytotoxic drug, a cytostatic drug or an immunomodulatory agent.

25. The antibody drug conjugate of claim 22 or 23, wherein the Drug Unit comprises an auristatin, a camptothecin, a calicheamycin or a maytansinoid.

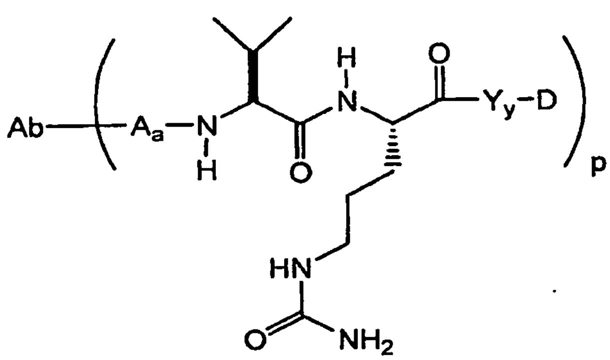
26. The antibody drug conjugate of claim 25, wherein the Drug Unit comprises an auristatin.

27. The antibody drug conjugate of claim 26, wherein the auristatin is MMAF or MMAE.

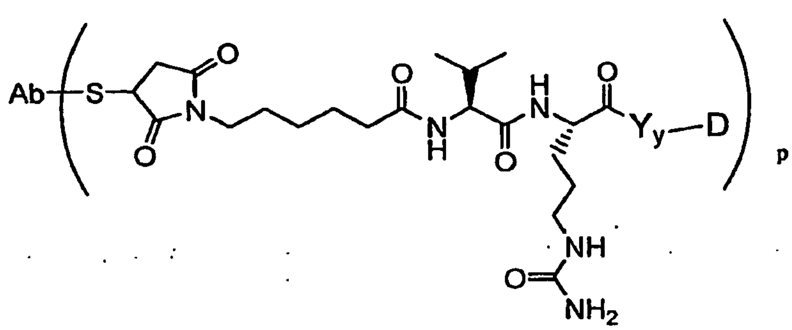
28. The antibody drug conjugate of claim 23, having the formula:



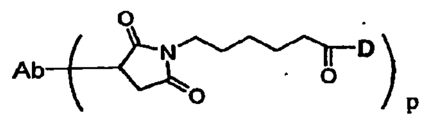
29. The antibody drug conjugate of claim 23, having the formula:



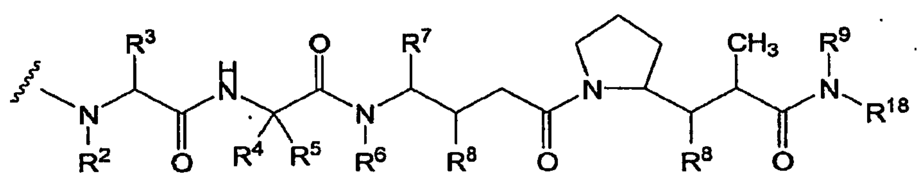
30. The antibody drug conjugate of claim 23, having the formula:



31. The antibody drug conjugate of claim 23, having the formula:

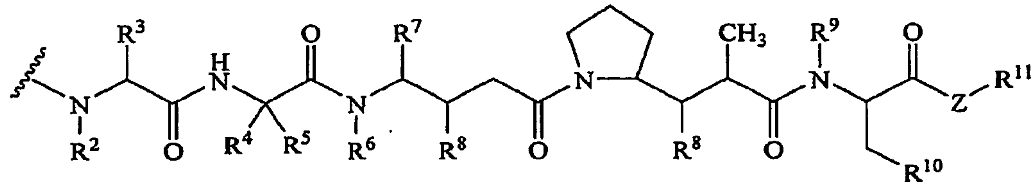


32. The antibody drug conjugate of claim 22 or 23, wherein the Drug Unit has the formula:



D_E

or



D_F

wherein, independently at each location:

R^2 is selected from H and C_1-C_8 alkyl;
R^3 is selected from H, C_1-C_8 alkyl, C_3-C_8 carbocycle, aryl, C_1-C_{10} alkyl-aryl, C_1-C_{10} alkyl-(C_3-C_8 carbocycle), C_3-C_8 heterocycle and C_1-C_{10} alkyl-(C_3-C_8 heterocycle);
R^4 is selected from H, C_1-C_8 alkyl, C_3-C_8 carbocycle, aryl, C_1-C_{10} alkyl-aryl, C_1-C_{10} alkyl-(C_3-C_8 carbocycle), C_3-C_8 heterocycle and C_1-C_{10} alkyl-(C_3-C_8 heterocycle);
R^5 is selected from H and methyl;
or R^4 and R^5 jointly form a carbocyclic ring and have the formula $-(CR^aR^b)_n-$ wherein R^a and R^b are independently selected from H, C_1-C_8 alkyl and C_3-C_8 carbocycle and n is selected from 2, 3, 4, 5 and 6;
R^6 is selected from H and C_1-C_8 alkyl;

R^7 is selected from H, C_1-C_8 alkyl, C_3-C_8 carbocycle, aryl, C_1-C_{10} alkyl-aryl, C_1-C_{10} alkyl-(C_3-C_8 carbocycle), C_3-C_8 heterocycle and C_1-C_{10} alkyl-(C_3-C_8 heterocycle);
each R^8 is independently selected from H, OH, C_1-C_8 alkyl, C_3-C_8 carbocycle and O-(C_1-C_8 alkyl);
R^9 is selected from H and C_1-C_8 alkyl;
R^{10} is selected from aryl or C_3-C_8 heterocycle;
Z is O, S, NH, or NR^{12}, wherein R^{12} is C_1-C_8 alkyl;
R^{11} is selected from H, C_1-C_{20} alkyl, aryl, C_3-C_8 heterocycle, $-(R^{13}O)_m-R^{14}$, or $-(R^{13}O)_m-CH(R^{15})_2$;
m is an integer ranging from 1-1000;
R^{13} is C_2-C_8 alkyl;
R^{14} is H or C_1-C_8 alkyl;
each occurrence of R^{15} is independently H, COOH, $-(CH_2)_n-N(R^{16})_2$, $-(CH_2)_n-SO_3H$, or $-(CH_2)_n-SO_3-C_1-C_8$ alkyl;
each occurrence of R^{16} is independently H, C_1-C_8 alkyl, or $-(CH_2)_n-COOH$;
R^{18} is selected from $-C(R^8)_2-C(R^8)_2-$aryl, $-C(R^8)_2-C(R^8)_2-(C_3-C_8$ heterocycle), and $-C(R^8)_2-C(R^8)_2-(C_3-C_8$ carbocycle); and
n is an integer ranging from 0 to 6;
or a pharmaceutically acceptable salt or solvate thereof.

33. The use of the antibody drug conjugate of claims 22-32 in the preparation of a medicament for the treatment of a CD133-expressing cancer.

Prominin-1 mRNA Expression Analysis: Cell Line Panel

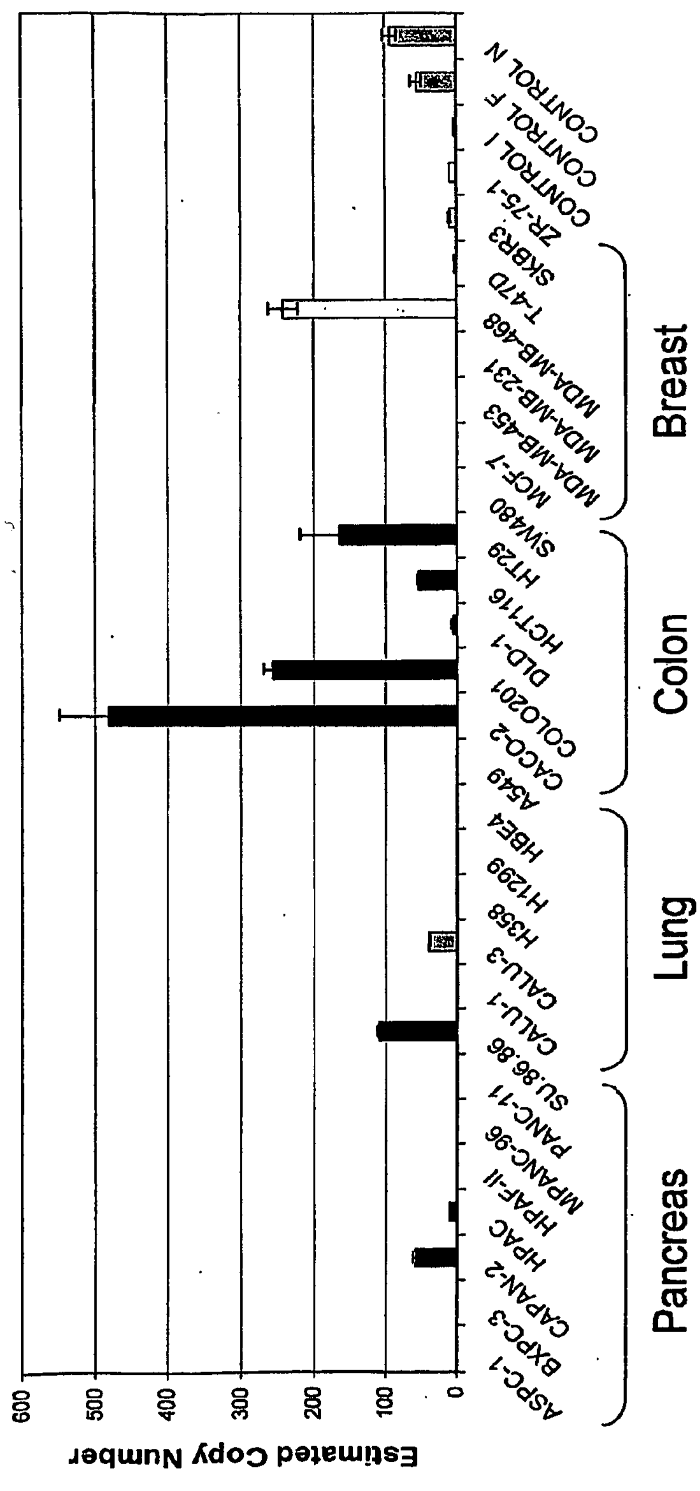


Figure 1

Prominin-1 mRNA Expression Analysis: Cell Line Panel

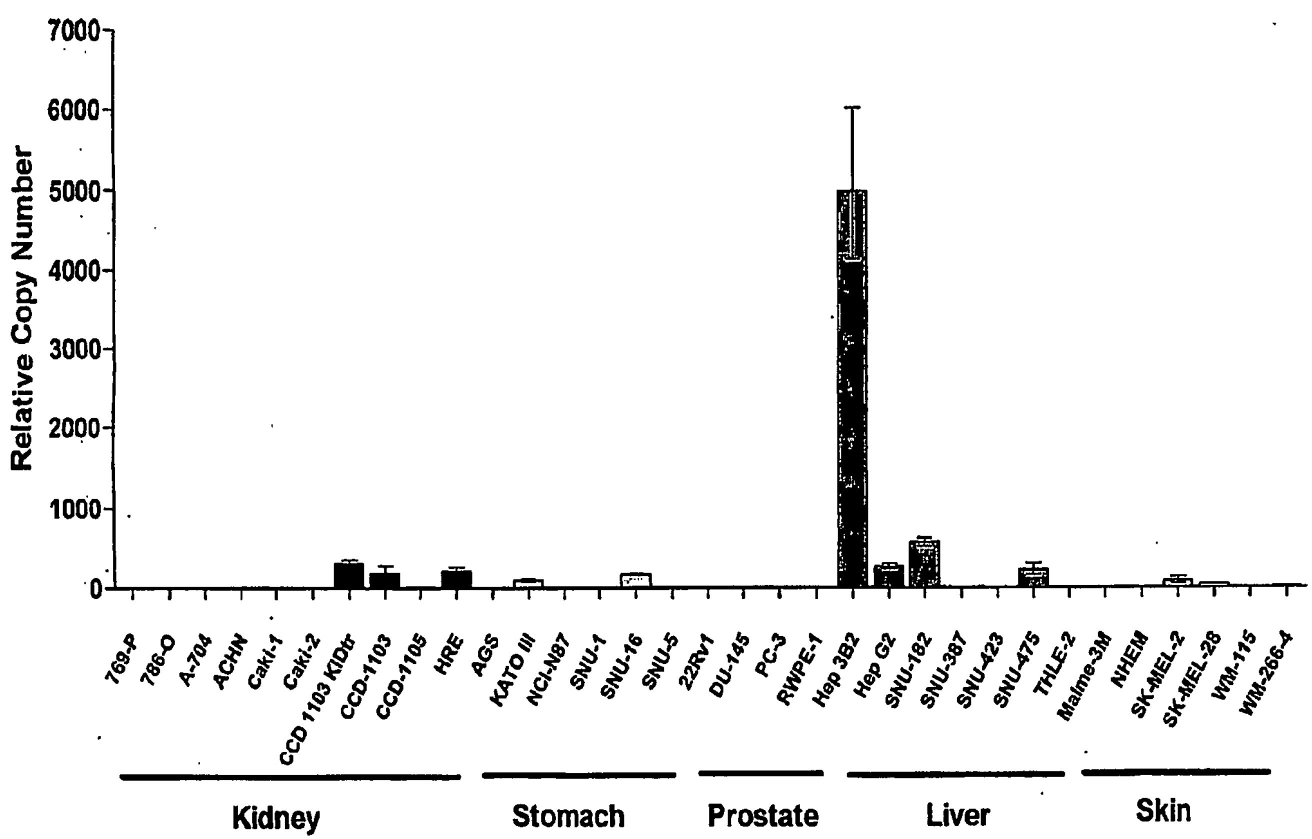


Figure 2

Prominin-1 is Overexpressed in Multiple Tumor Types

	Overexpression Δ+1 (% tumors)	Overexpression Δ+2 (% tumors)
Kidney	70%	10%
Pancreas, Metastatic	67%	67%
Colon	60%	60%
Pharynageal	60%	60%
Liver	50%	50%
Melanoma	50%	30%
Bladder	50%	0%
Lung, NSC	40%	40%
Ovary	40%	40%
Prostate	40%	40%
Lung, Squamous	30%	30%
Breast	30%	0%

Figure 3

Figure 4

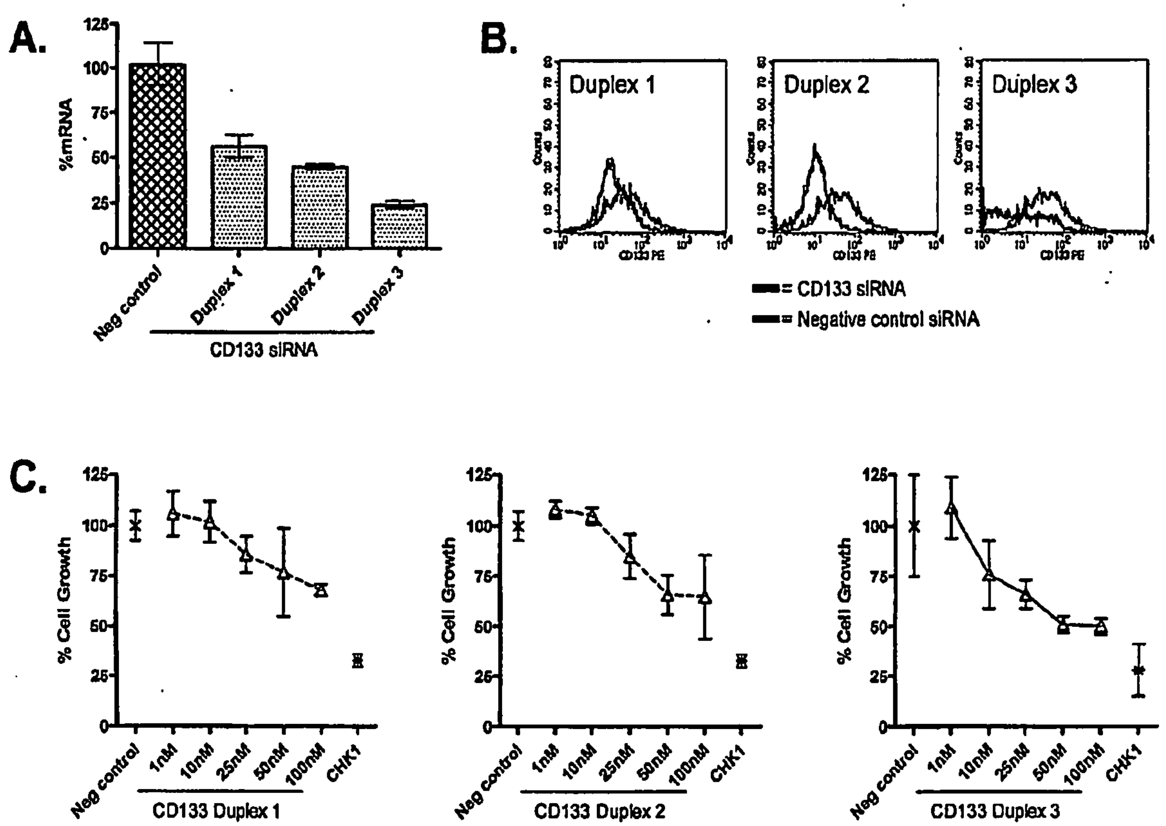

A.
%mRNA
125
100
75
50
25
0
Neg control
Duplex 1
Duplex 2
Duplex 3
CD133 siRNA
B.
Duplex 1
Duplex 2
Duplex 3
= CD133 siRNA
= Negative control siRNA
C.
% Cell Growth
Neg control
1nM
10nM
25nM
50nM
100nM
CHK1
CD133 Duplex 1
CD133 Duplex 2
CD133 Duplex 3

FIGURE 5

mRNA Sequence of Prominin-1 (CD133)

ccaagttctacctcatgtttggaggatcttgctagctatggccctcgtactcggctccctgttgctgctggggctgtgcgggaactccttttcaggagggcagccttcatccacagatgctcctaag
gcttggaattatgaattgcctgcaacaaattatgagacccaagactcccataaagctggacccattggcattctctttgaactagtgcatatctttctctatgtggtacagccgcgtgatttcccag
aagatactttgagaaaattcttacagaaggcatatgaatccaaaattgattatgacaag**ccagaaactgtaatctta**ggtctaaagattgtctactatgaagcagggattattctatgct
gtgtcctggggctgctgtttattattctgatgcctctggtggggtatttcttttgtatgtgtcgttgctg**taacaaatgtggtggagaa**atgcaccagcgacagaaggaaaatgggcccttcc
tgaggaaatgctttgcaatctccctgttggtgatttgtataataataagcattggcatcttctatggttttgtggcaaatcaccaggtaagaacccggatcaaaaggagtcggaaactggcagat
agcaatttcaaggacttgcgaactctcttgaatgaaactccagagcaaatcaaatatatattggcccagtacaacactaccaaggacaaggcgttcacagatctgaacagtatcaattcag
tgctaggaggcggaattcttgaccgactgagacccaacatcatccctgttcttgatgagattaagtccatggcaacagcgatcaaggagaccaaagaggcgttggagaacatgaacagc
accttgaagagcttgcaccaacaaagtacacagcttagcagcagtctgaccagcgtgaaaactagcctgcggtcatctctcaatgaccctctgtgcttggtgcatccatcaagtgaaacctg
caacagcatcagattgtctctaagccagctgaatagcaaccctgaactgaggcagcttccacccgtggatgcagaacttgacaacgttaataacgttcttaggacagatttggatggcctgg
tccaacagggctatcaatcccttaatgatatacctgacagagtacaacgccaaaccacgactgtcgtagcaggtatcaaaagggtcttgaattccattggttcagatatcgacaatgtaactc
agcgtcttcctattcaggatatactctcagcattctctgtttatgttaataacactgaaagttacatccacagaaatttacctacattggaagagtatgattcatactggtggctgggtggcctggtc
atctgctctctgctgaccctcatcgtgatttttactacctgggcttactgtgtggcgtgtgcggctatgacaggcatgccacccccgaccacccgaggctgtgtctccaacaccggaggcgtcttc
ctcatggttggagttggattaagtttcctcttttgctggatattgatgatcattgtggttcttacctttgtctttggtgcaaatgtggaaaaactgatctgtgaaccttacacgagcaaggaattattccg
ggttttggatacaccctacttactaaatgaagactgggaatactatctctctgggaagctatttaataaatcaaaaatgaagctcacttttgaacaagtttacagtgactgcaaaaaaaataga
ggcacttacggcactcttcacctgcagaacagcttcaatatcagtgaacatctcaacattaatgagcatactggaagcataagcagtgaattggaaagtctgaaggtaaatcttaatatctttc
tgttgggtgcagcaggaagaaaaaaccttcaggattttgctgcttgtggaatagacagaatgaattatgacagctacttggctcagactggtaaatcccccgcaggagtgaatcttttatcattt
gcatatgatctagaagcaaaagcaaacagtttgcccccaggaaatttgaggaactccctgaaaagagatgcacaaactattaaaacaattcaccagcaacgagtccttcctatagaaca
atcactgagcactctataccaaagcgtcaagatacttcaacgcacagggaatggattgttggagagagtaactaggattctagcttctctggatttt**gctcagaacttcatcacaa**ac
aatacttcctctgttattattgaggaaactaagaagtatgggagaacaataataggatattttgaacattatctgcagtggatcgagttctctatcagtgagaaagtggcatcgtgcaaacctgt
ggccaccgctctagatactgctgttgatgtctttctgtgtagctacattatcgaccccttgaatttgttttggtttggcataggaaaagctactgtatttttacttccggctctaattttgcggtaaaactg
gctaagtactatcgtcgaatggattcggaggacgtgtacgatgatgttgaaactatacccatgaaaaatatggaaaatggtaataatggttatcataaagatcatgtatatggtattcacaatc
ctgttatgacaagcccatcacaacattgatagctgatgttgaaactgcttgagcatcaggatactcaaagtggaaaggatcacagatttttggtagtttctgggtctacaaggactttccaaatc
caggagcaacgccagtggcaacgtagtgactcaggcgggcaccaaggcaacggcaccattggtctctgggtagtgctttaagaatgaacacaatcacgttatagtccatggtccatcact
attcaaggatgactccctcccttcctgtctattttgtttttactttttacactgagtttctatttagacactacaacatatggggtgtttgttcccattggatgcatttctatcaaaactctatcaaatgtgat
ggctagattctaacatattgccatgtgtggagtgtgctgaacacacaccagtttacaggaaagatgcattttgtgtacagtaaacggtgtatataccttttgttaccacagagtttttaaacaaat
gagtattataggactttcttctaaatgagctaaataagtcaccattgacttcttggtgctgttgaaaataatccattttcactaaaagtgtgtgaaacctacagcatattcttcacgcagagattttcat
ctattatactttatcaaagattggccatgttccacttggaaatggcatgcaaaagccatcatagagaaacctgcgtaactccatctgacaaattcaaaagagagagagagatcttgagaga
gaaatgctgttcgttcaaaagtggagttgttttaacagatgccaattacggtgtacagtttaacagagttttctgttgcattaggataaacattaattggagtgcagctaacatgagtatcatcaga
ctagtatcaagtgttctaaaatgaaatatgagaagatcctgtcacaattcttagatctggtgtccagcatggatgaaacctttgagtttggtccctaaatttgcatgaaagcacaaggtaaatatt
catttgcttcaggagtttcatgttggatctgtcattatcaaaagtgatcagcaatgaagaactggtcggacaaaatttaacgttgatgtaatggaattccagatgtaggcattcccccaggtctttt
catgtgcagattgcagttctgattcatttgaataaaaaggaacttgg (SEQ ID NO:11)

Bold underlined indicates siRNA target region with dose-dependent biological activity

Figure 6A

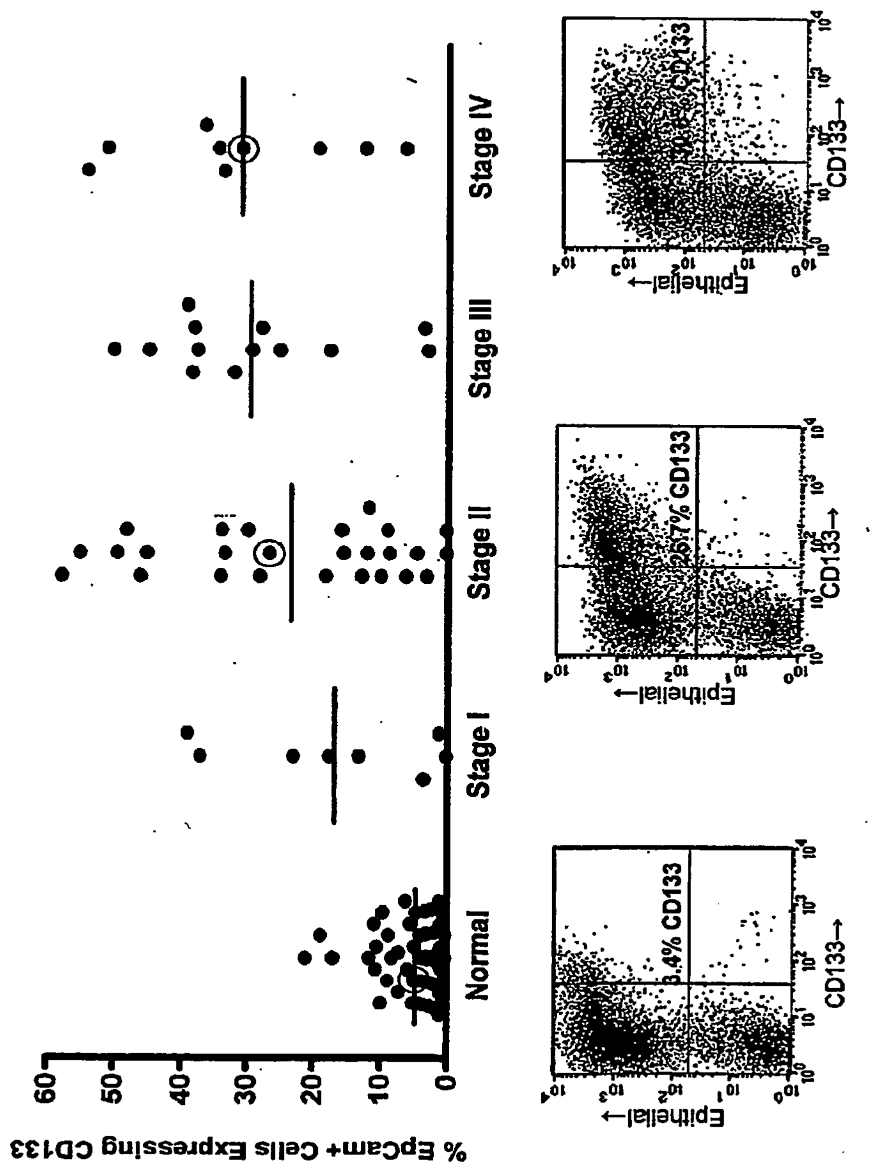
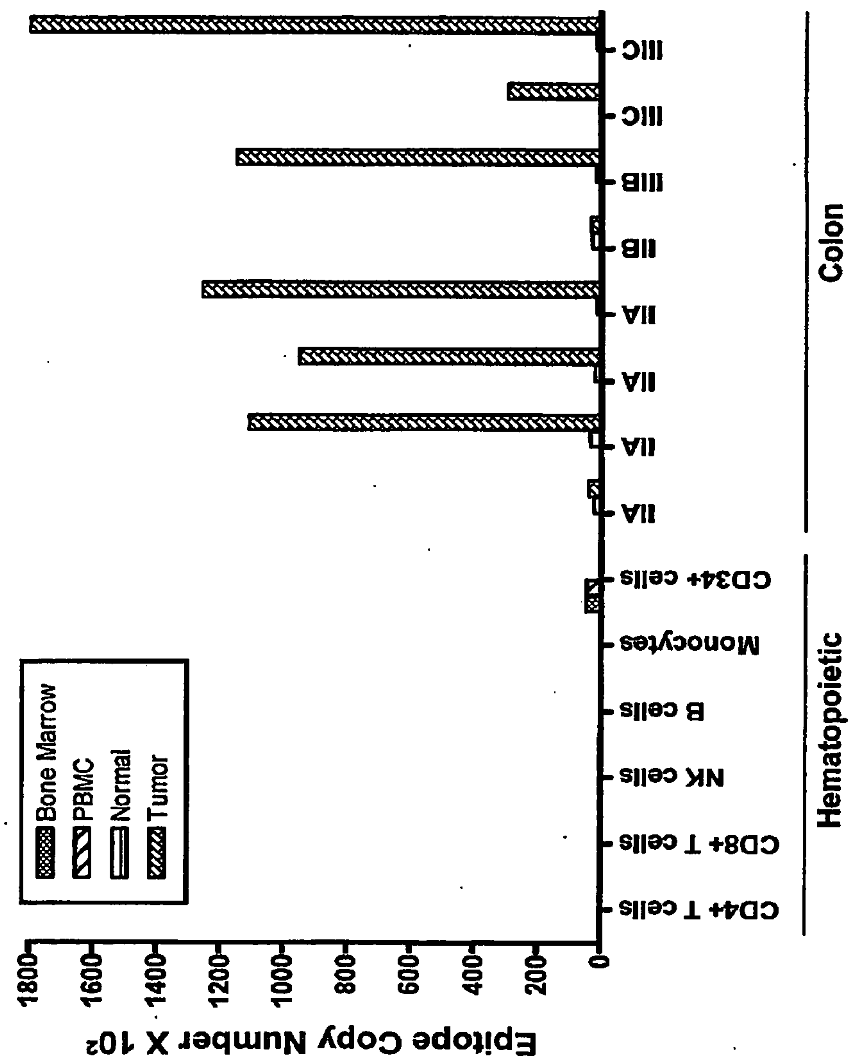

% EpCam+ Cells Expressing CD133
60
50
40
30
20
10
0
Normal
Stage I
Stage II
Stage III
Stage IV
Epithelial→
CD133→
3.4% CD133
CD133→
Epithelial→

Figure 6B


Epitope Copy Number X 10²
0
200
400
600
800
1000
1200
1400
1600
1800
Bone Marrow
PBMC
Normal
Tumor
CD4+ T cells
CD8+ T cells
NK cells
B cells
Monocytes
CD34+ cells
IIA
IIA
IIA
IIA
IIB
IIIB
IIIC
IIIC
Hematopoietic
Colon

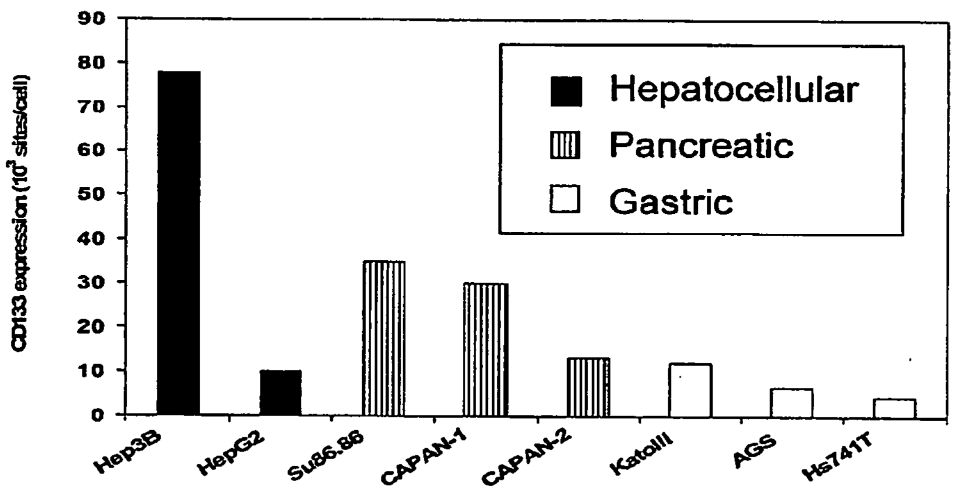
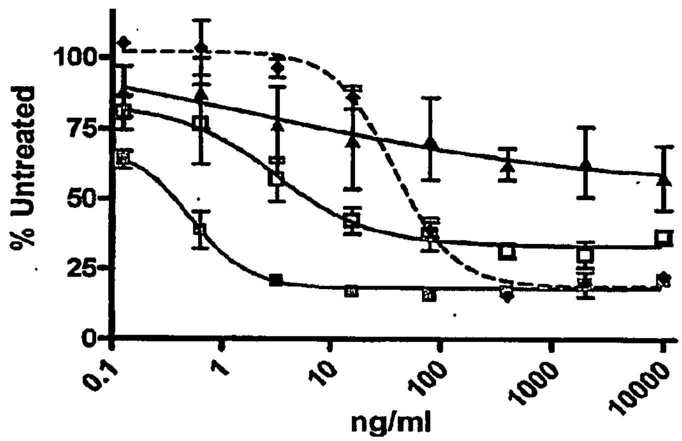
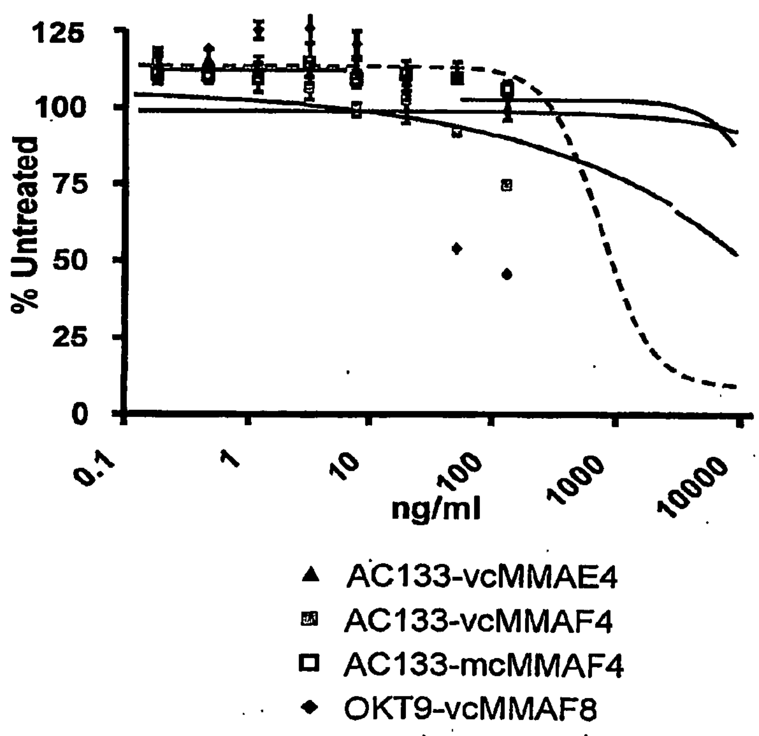

90
80
70
60
50
40
30
20
10
0
CD133 expression (10^3 sites/cell)
Hepatocellular
Pancreatic
Gastric
Hep3B
HepG2
Su86.86
CAPAN-1
CAPAN-2
KatoIII
AGS
Hs741T

Figure 7A

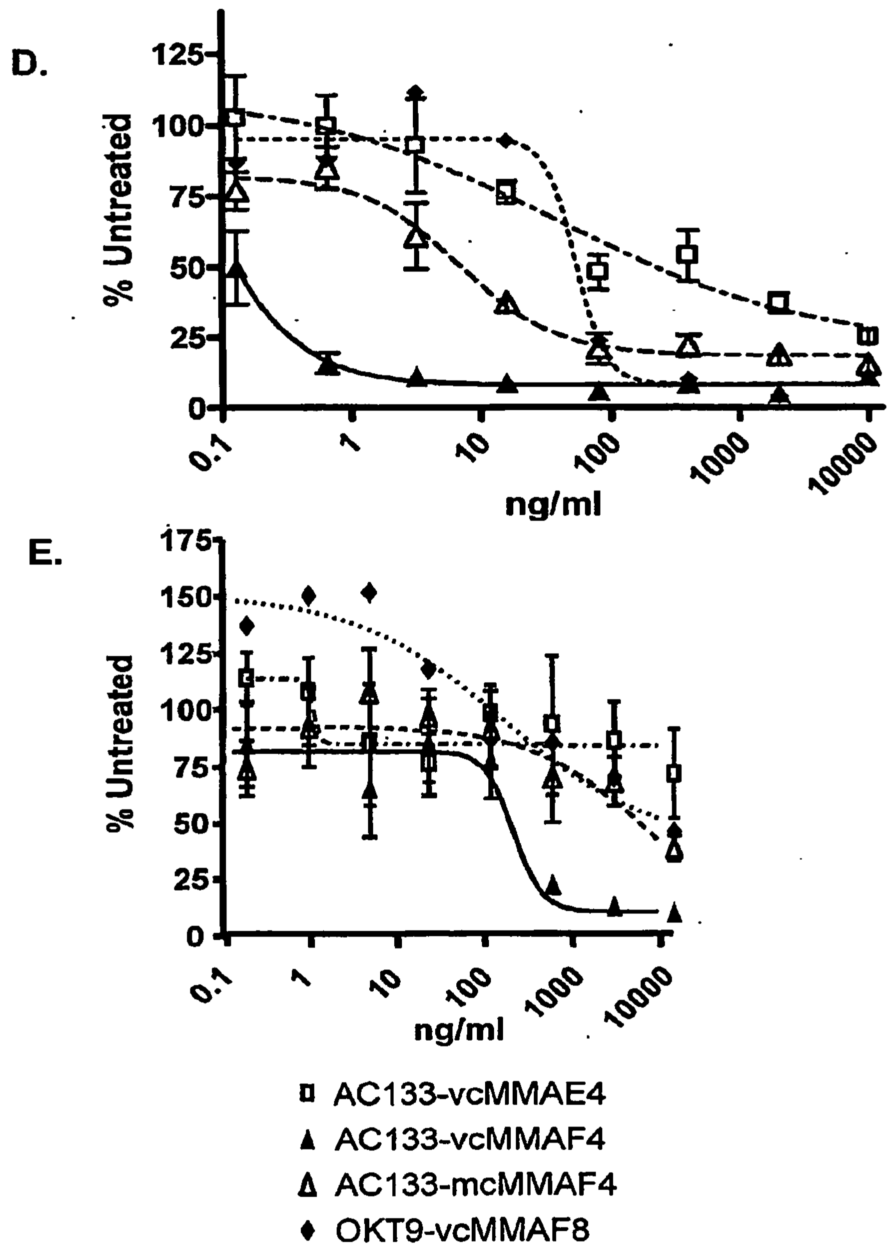
B.



C.



Figures 7B-7C



Figures 7D-7E

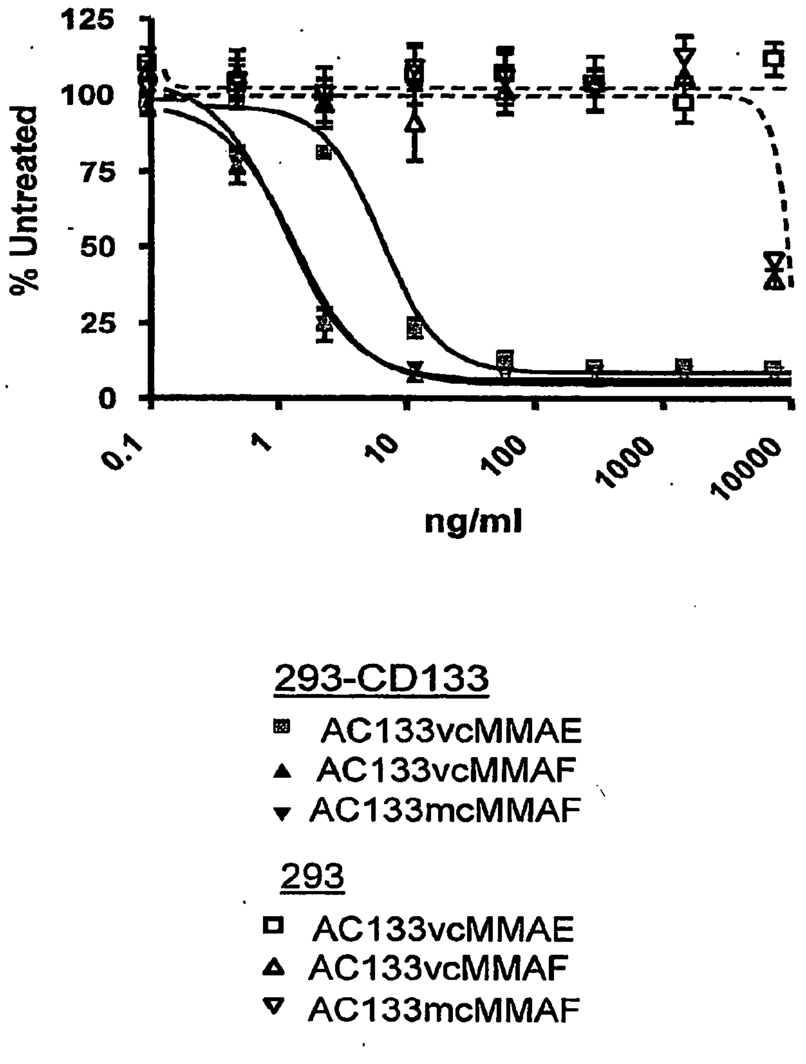

125
100
75
50
25
0
% Untreated
0.1
1
10
100
1000
10000
ng/ml
293-CD133
AC133vcMMAE
AC133vcMMAF
AC133mcMMAF
293
AC133vcMMAE
AC133vcMMAF
AC133mcMMAF

Figure 8A

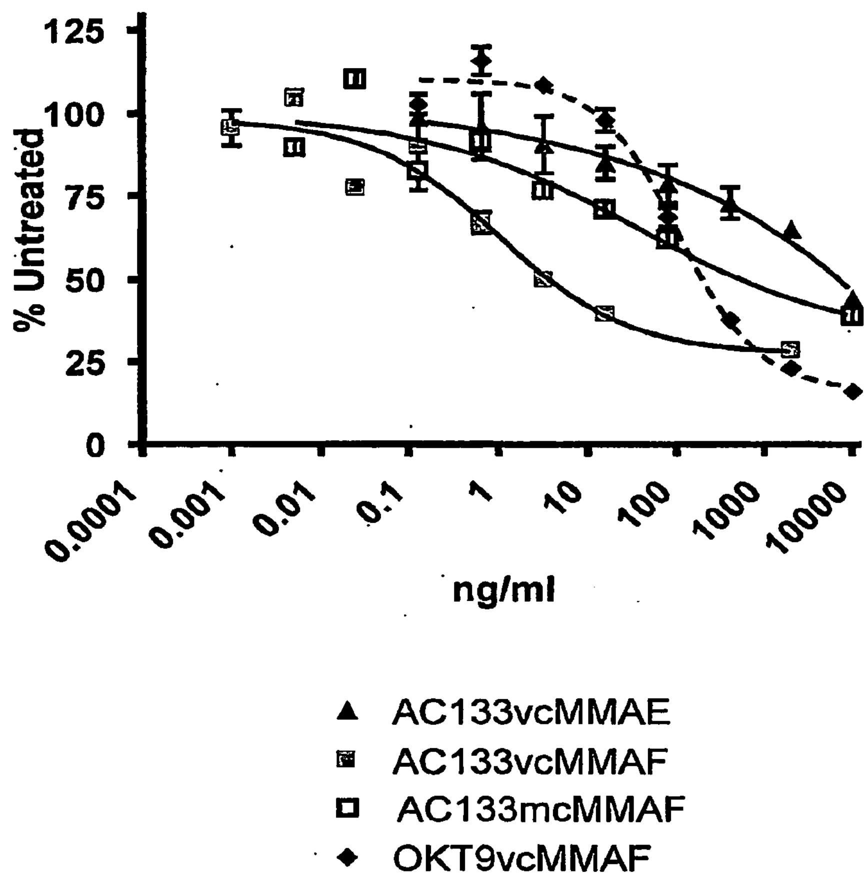

125
100
75
50
25
0
% Untreated
0.0001
0.001
0.01
0.1
1
10
100
1000
10000
ng/ml
AC133vcMMAE
AC133vcMMAF
AC133mcMMAF
OKT9vcMMAF

Figure 8B

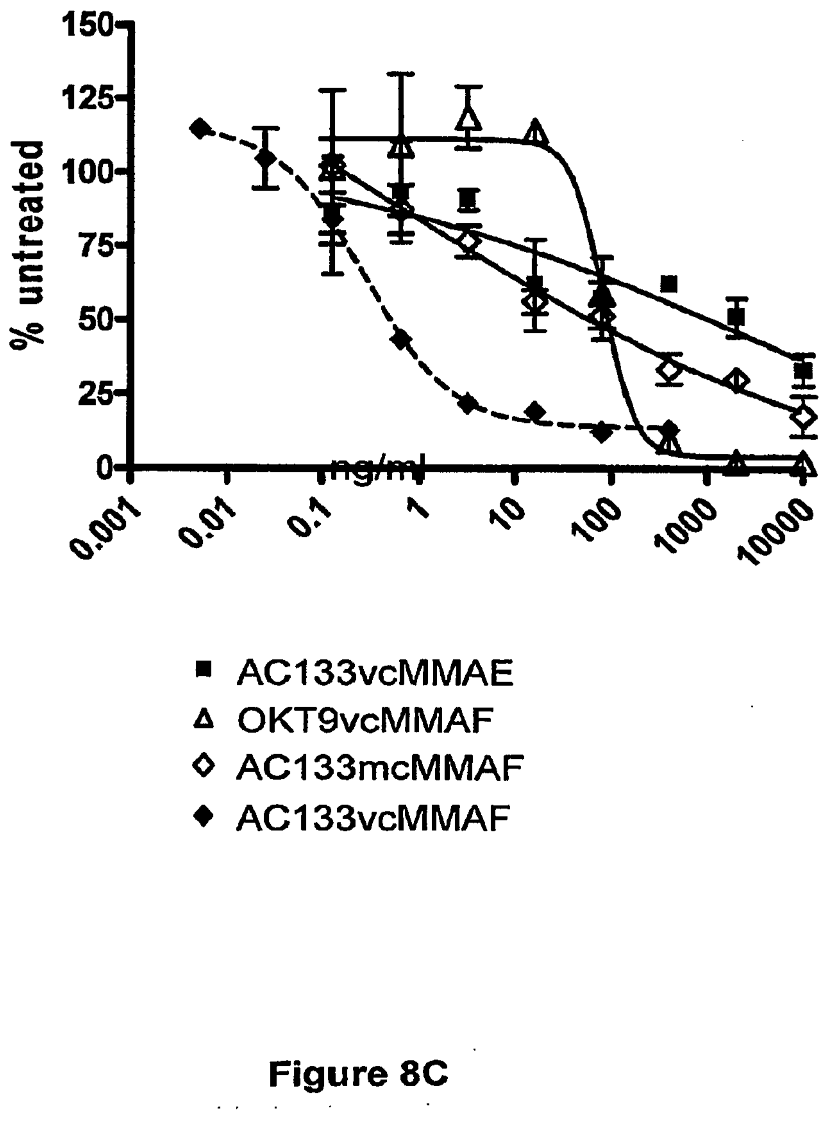

150
125
100
75
50
25
0
% untreated
ng/ml
0.001
0.01
0.1
1
10
100
1000
10000
AC133vcMMAE
OKT9vcMMAF
AC133mcMMAF
AC133vcMMAF

Figure 8C

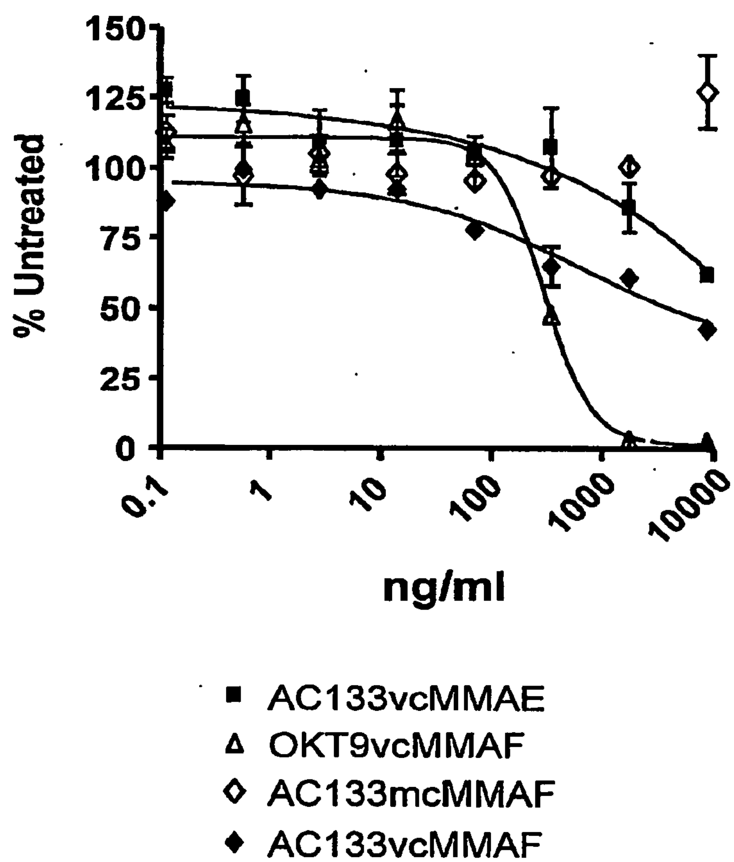

150
125
100
75
50
25
0
% Untreated
0.1
1
10
100
1000
10000
ng/ml
AC133vcMMAE
OKT9vcMMAF
AC133mcMMAF
AC133vcMMAF

Figure 8D

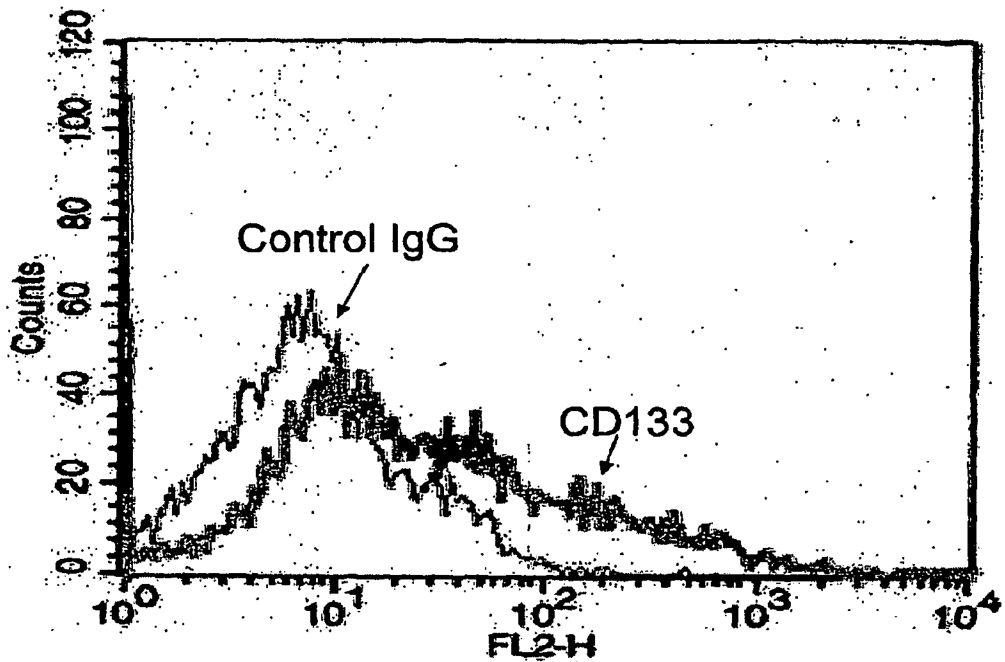
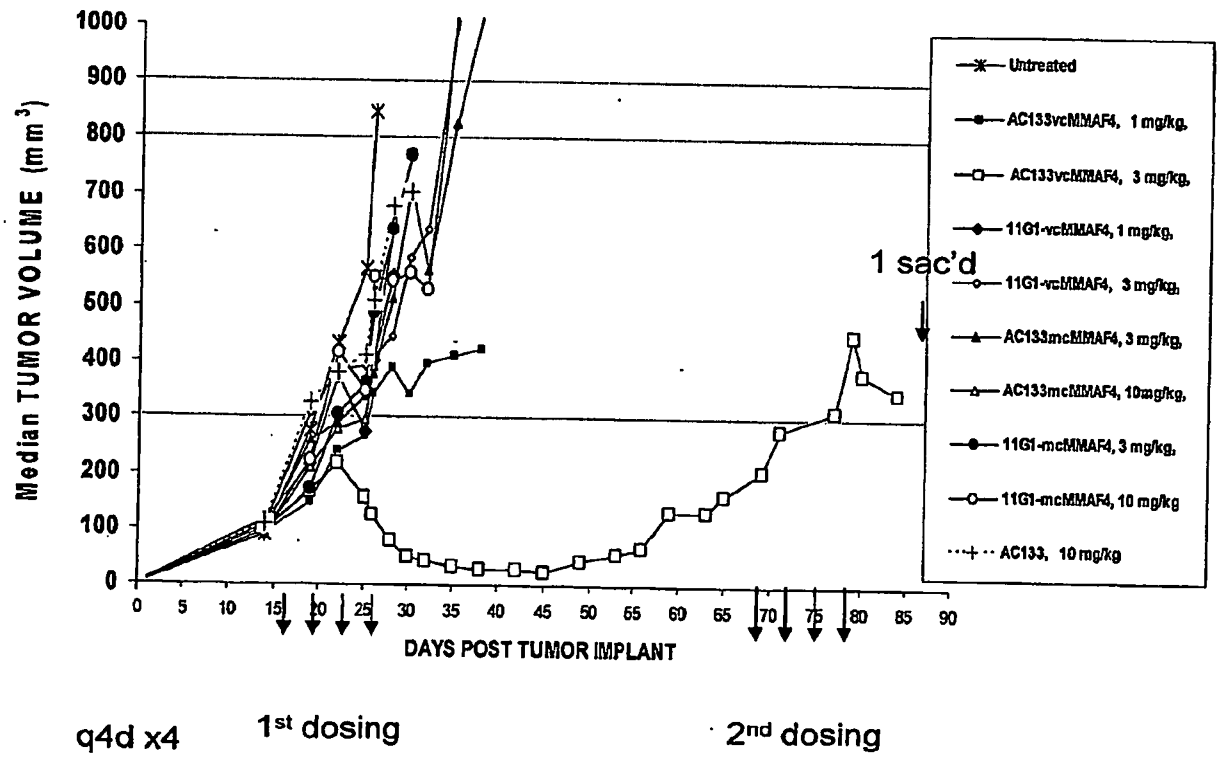

Control IgG
CD133
Counts
FL2-H
0
20
40
60
80
100
120
10^0
10^1
10^2
10^3
10^4

Figure 9A



1000
900
800
700
600
500
400
300
200
100
0
Median TUMOR VOLUME (mm³)
0 5 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90
DAYS POST TUMOR IMPLANT
1 sac'd
Untreated
AC133vcMMAF4, 1 mg/kg,
AC133vcMMAF4, 3 mg/kg,
11G1-vcMMAF4, 1 mg/kg,
11G1-vcMMAF4, 3 mg/kg,
AC133mcMMAF4, 3 mg/kg,
AC133mcMMAF4, 10mg/kg,
11G1-mcMMAF4, 3 mg/kg,
11G1-mcMMAF4, 10 mg/kg
AC133, 10 mg/kg
q4d x4
1st dosing
2nd dosing

Figure 9B

FIGURE 10
Cell Transformation Assay
Soft Agar Colony Formation

- Cells grown in 0.4% agar and colonies counted
- Measures tumorigenic potential of CD133

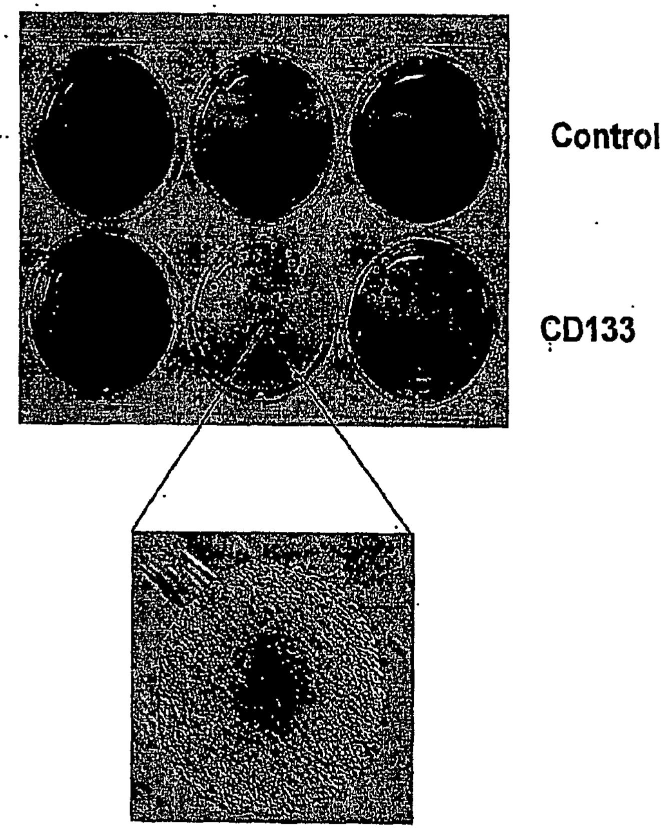


TABLE 1

Tissue Sample	Number of Cases	CD133 + *	CD133-
Total	29	22 (76%)	7 (27%)
Undifferentiated/Poor	8	5	3
Moderate/ Well-Differentiated	21	17	4

SEQUENCE LISTING

<110> APPLERA CORPORATION et al.

<120> METHODS AND COMPOSITIONS FOR TREATING
DISEASES TARGETING PROMININ-1 (CD133)

<130> CL001639-PCT

<150> 60/738,965
<151> 2005-11-23

<160> 11

<170> FastSEQ for Windows Version 4.0

<210> 1
<211> 865
<212> PRT
<213> Homo sapiens

<400> 1

```
Met Ala Leu Val Leu Gly Ser Leu Leu Leu Leu Gly Leu Cys Gly Asn
 1               5                  10                  15
Ser Phe Ser Gly Gly Gln Pro Ser Ser Thr Asp Ala Pro Lys Ala Trp
            20                  25                  30
Asn Tyr Glu Leu Pro Ala Thr Asn Tyr Glu Thr Gln Asp Ser His Lys
        35                  40                  45
Ala Gly Pro Ile Gly Ile Leu Phe Glu Leu Val His Ile Phe Leu Tyr
    50                  55                  60
Val Val Gln Pro Arg Asp Phe Pro Glu Asp Thr Leu Arg Lys Phe Leu
65                  70                  75                  80
Gln Lys Ala Tyr Glu Ser Lys Ile Asp Tyr Asp Lys Pro Glu Thr Val
                85                  90                  95
Ile Leu Gly Leu Lys Ile Val Tyr Tyr Glu Ala Gly Ile Ile Leu Cys
            100                 105                 110
Cys Val Leu Gly Leu Leu Phe Ile Ile Leu Met Pro Leu Val Gly Tyr
        115                 120                 125
Phe Phe Cys Met Cys Arg Cys Cys Asn Lys Cys Gly Gly Glu Met His
    130                 135                 140
Gln Arg Gln Lys Glu Asn Gly Pro Phe Leu Arg Lys Cys Phe Ala Ile
145                 150                 155                 160
Ser Leu Leu Val Ile Cys Ile Ile Ile Ser Ile Gly Ile Phe Tyr Gly
                165                 170                 175
Phe Val Ala Asn His Gln Val Arg Thr Arg Ile Lys Arg Ser Arg Lys
            180                 185                 190
Leu Ala Asp Ser Asn Phe Lys Asp Leu Arg Thr Leu Leu Asn Glu Thr
        195                 200                 205
Pro Glu Gln Ile Lys Tyr Ile Leu Ala Gln Tyr Asn Thr Thr Lys Asp
    210                 215                 220
Lys Ala Phe Thr Asp Leu Asn Ser Ile Asn Ser Val Leu Gly Gly Gly
225                 230                 235                 240
Ile Leu Asp Arg Leu Arg Pro Asn Ile Ile Pro Val Leu Asp Glu Ile
                245                 250                 255
Lys Ser Met Ala Thr Ala Ile Lys Glu Thr Lys Glu Ala Leu Glu Asn
```

```
            260                 265                 270
Met Asn Ser Thr Leu Lys Ser Leu His Gln Gln Ser Thr Gln Leu Ser
        275                 280                 285
Ser Ser Leu Thr Ser Val Lys Thr Ser Leu Arg Ser Ser Leu Asn Asp
    290                 295                 300
Pro Leu Cys Leu Val His Pro Ser Ser Glu Thr Cys Asn Ser Ile Arg
305                 310                 315                 320
Leu Ser Leu Ser Gln Leu Asn Ser Asn Pro Glu Leu Arg Gln Leu Pro
                325                 330                 335
Pro Val Asp Ala Glu Leu Asp Asn Val Asn Asn Val Leu Arg Thr Asp
            340                 345                 350
Leu Asp Gly Leu Val Gln Gln Gly Tyr Gln Ser Leu Asn Asp Ile Pro
        355                 360                 365
Asp Arg Val Gln Arg Gln Thr Thr Thr Val Val Ala Gly Ile Lys Arg
    370                 375                 380
Val Leu Asn Ser Ile Gly Ser Asp Ile Asp Asn Val Thr Gln Arg Leu
385                 390                 395                 400
Pro Ile Gln Asp Ile Leu Ser Ala Phe Ser Val Tyr Val Asn Asn Thr
                405                 410                 415
Glu Ser Tyr Ile His Arg Asn Leu Pro Thr Leu Glu Glu Tyr Asp Ser
            420                 425                 430
Tyr Trp Trp Leu Gly Gly Leu Val Ile Cys Ser Leu Leu Thr Leu Ile
        435                 440                 445
Val Ile Phe Tyr Tyr Leu Gly Leu Leu Cys Gly Val Cys Gly Tyr Asp
    450                 455                 460
Arg His Ala Thr Pro Thr Thr Arg Gly Cys Val Ser Asn Thr Gly Gly
465                 470                 475                 480
Val Phe Leu Met Val Gly Val Gly Leu Ser Phe Leu Phe Cys Trp Ile
                485                 490                 495
Leu Met Ile Ile Val Val Leu Thr Phe Val Phe Gly Ala Asn Val Glu
            500                 505                 510
Lys Leu Ile Cys Glu Pro Tyr Thr Ser Lys Glu Leu Phe Arg Val Leu
        515                 520                 525
Asp Thr Pro Tyr Leu Leu Asn Glu Asp Trp Glu Tyr Tyr Leu Ser Gly
    530                 535                 540
Lys Leu Phe Asn Lys Ser Lys Met Lys Leu Thr Phe Glu Gln Val Tyr
545                 550                 555                 560
Ser Asp Cys Lys Lys Asn Arg Gly Thr Tyr Gly Thr Leu His Leu Gln
                565                 570                 575
Asn Ser Phe Asn Ile Ser Glu His Leu Asn Ile Asn Glu His Thr Gly
            580                 585                 590
Ser Ile Ser Ser Glu Leu Glu Ser Leu Lys Val Asn Leu Asn Ile Phe
        595                 600                 605
Leu Leu Gly Ala Ala Gly Arg Lys Asn Leu Gln Asp Phe Ala Ala Cys
    610                 615                 620
Gly Ile Asp Arg Met Asn Tyr Asp Ser Tyr Leu Ala Gln Thr Gly Lys
625                 630                 635                 640
Ser Pro Ala Gly Val Asn Leu Leu Ser Phe Ala Tyr Asp Leu Glu Ala
                645                 650                 655
Lys Ala Asn Ser Leu Pro Pro Gly Asn Leu Arg Asn Ser Leu Lys Arg
            660                 665                 670
Asp Ala Gln Thr Ile Lys Thr Ile His Gln Gln Arg Val Leu Pro Ile
        675                 680                 685
Glu Gln Ser Leu Ser Thr Leu Tyr Gln Ser Val Lys Ile Leu Gln Arg
    690                 695                 700
Thr Gly Asn Gly Leu Leu Glu Arg Val Thr Arg Ile Leu Ala Ser Leu
705                 710                 715                 720
```

```
Asp Phe Ala Gln Asn Phe Ile Thr Asn Asn Thr Ser Ser Val Ile Ile
                725                 730                 735
Glu Glu Thr Lys Lys Tyr Gly Arg Thr Ile Ile Gly Tyr Phe Glu His
            740                 745                 750
Tyr Leu Gln Trp Ile Glu Phe Ser Ile Ser Glu Lys Val Ala Ser Cys
        755                 760                 765
Lys Pro Val Ala Thr Ala Leu Asp Thr Ala Val Asp Val Phe Leu Cys
    770                 775                 780
Ser Tyr Ile Ile Asp Pro Leu Asn Leu Phe Trp Phe Gly Ile Gly Lys
785                 790                 795                 800
Ala Thr Val Phe Leu Leu Pro Ala Leu Ile Phe Ala Val Lys Leu Ala
                805                 810                 815
Lys Tyr Tyr Arg Arg Met Asp Ser Glu Asp Val Tyr Asp Asp Val Glu
            820                 825                 830
Thr Ile Pro Met Lys Asn Met Glu Asn Gly Asn Asn Gly Tyr His Lys
        835                 840                 845
Asp His Val Tyr Gly Ile His Asn Pro Val Met Thr Ser Pro Ser Gln
    850                 855                 860
His
865


<210> 2
<211> 856
<212> PRT
<213> Homo sapiens

<400> 2
Met Ala Leu Val Leu Gly Ser Leu Leu Leu Leu Gly Leu Cys Gly Asn
 1               5                  10                  15
Ser Phe Ser Gly Gly Gln Pro Ser Ser Thr Asp Ala Pro Lys Ala Trp
            20                  25                  30
Asn Tyr Glu Leu Pro Ala Thr Asn Tyr Glu Thr Gln Asp Ser His Lys
        35                  40                  45
Ala Gly Pro Ile Gly Ile Leu Phe Glu Leu Val His Ile Phe Leu Tyr
    50                  55                  60
Val Val Gln Pro Arg Asp Phe Pro Glu Asp Thr Leu Arg Lys Phe Leu
65                  70                  75                  80
Gln Lys Ala Tyr Glu Ser Lys Ile Asp Tyr Asp Lys Ile Val Tyr Tyr
                85                  90                  95
Glu Ala Gly Ile Ile Leu Cys Cys Val Leu Gly Leu Leu Phe Ile Ile
            100                 105                 110
Leu Met Pro Leu Val Gly Tyr Phe Phe Cys Met Cys Arg Cys Cys Asn
        115                 120                 125
Lys Cys Gly Gly Glu Met His Gln Arg Gln Lys Glu Asn Gly Pro Phe
    130                 135                 140
Leu Arg Lys Cys Phe Ala Ile Ser Leu Leu Val Ile Cys Ile Ile Ile
145                 150                 155                 160
Ser Ile Gly Ile Phe Tyr Gly Phe Val Ala Asn His Gln Val Arg Thr
                165                 170                 175
Arg Ile Lys Arg Ser Arg Lys Leu Ala Asp Ser Asn Phe Lys Asp Leu
            180                 185                 190
Arg Thr Leu Leu Asn Glu Thr Pro Glu Gln Ile Lys Tyr Ile Leu Ala
        195                 200                 205
Gln Tyr Asn Thr Thr Lys Asp Lys Ala Phe Thr Asp Leu Asn Ser Ile
    210                 215                 220
Asn Ser Val Leu Gly Gly Gly Ile Leu Asp Arg Leu Arg Pro Asn Ile
```

```
225                 230                 235                 240
Ile Pro Val Leu Asp Glu Ile Lys Ser Met Ala Thr Ala Ile Lys Glu
                245                 250                 255
Thr Lys Glu Ala Leu Glu Asn Met Asn Ser Thr Leu Lys Ser Leu His
            260                 265                 270
Gln Gln Ser Thr Gln Leu Ser Ser Ser Leu Thr Ser Val Lys Thr Ser
        275                 280                 285
Leu Arg Ser Ser Leu Asn Asp Pro Leu Cys Leu Val His Pro Ser Ser
    290                 295                 300
Glu Thr Cys Asn Ser Ile Arg Leu Ser Leu Ser Gln Leu Asn Ser Asn
305                 310                 315                 320
Pro Glu Leu Arg Gln Leu Pro Pro Val Asp Ala Glu Leu Asp Asn Val
                325                 330                 335
Asn Asn Val Leu Arg Thr Asp Leu Asp Gly Leu Val Gln Gln Gly Tyr
            340                 345                 350
Gln Ser Leu Asn Asp Ile Pro Asp Arg Val Gln Arg Gln Thr Thr Thr
        355                 360                 365
Val Val Ala Gly Ile Lys Arg Val Leu Asn Ser Ile Gly Ser Asp Ile
    370                 375                 380
Asp Asn Val Thr Gln Arg Leu Pro Ile Gln Asp Ile Leu Ser Ala Phe
385                 390                 395                 400
Ser Val Tyr Val Asn Asn Thr Glu Ser Tyr Ile His Arg Asn Leu Pro
                405                 410                 415
Thr Leu Glu Glu Tyr Asp Ser Tyr Trp Trp Leu Gly Gly Leu Val Ile
            420                 425                 430
Cys Ser Leu Leu Thr Leu Ile Val Ile Phe Tyr Tyr Leu Gly Leu Leu
        435                 440                 445
Cys Gly Val Cys Gly Tyr Asp Arg His Ala Thr Pro Thr Thr Arg Gly
    450                 455                 460
Cys Val Ser Asn Thr Gly Gly Val Phe Leu Met Val Gly Val Gly Leu
465                 470                 475                 480
Ser Phe Leu Phe Cys Trp Ile Leu Met Ile Ile Val Val Leu Thr Phe
                485                 490                 495
Val Phe Gly Ala Asn Val Glu Lys Leu Ile Cys Glu Pro Tyr Thr Ser
            500                 505                 510
Lys Glu Leu Phe Arg Val Leu Asp Thr Pro Tyr Leu Leu Asn Glu Asp
        515                 520                 525
Trp Glu Tyr Tyr Leu Ser Gly Lys Leu Phe Asn Lys Ser Lys Met Lys
    530                 535                 540
Leu Thr Phe Glu Gln Val Tyr Ser Asp Cys Lys Lys Asn Arg Gly Thr
545                 550                 555                 560
Tyr Gly Thr Leu His Leu Gln Asn Ser Phe Asn Ile Ser Glu His Leu
                565                 570                 575
Asn Ile Asn Glu His Thr Gly Ser Ile Ser Ser Glu Leu Glu Ser Leu
            580                 585                 590
Lys Val Asn Leu Asn Ile Phe Leu Leu Gly Ala Ala Gly Arg Lys Asn
        595                 600                 605
Leu Gln Asp Phe Ala Ala Cys Gly Ile Asp Arg Met Asn Tyr Asp Ser
    610                 615                 620
Tyr Leu Ala Gln Thr Gly Lys Ser Pro Ala Gly Val Asn Leu Leu Ser
625                 630                 635                 640
Phe Ala Tyr Asp Leu Glu Ala Lys Ala Asn Ser Leu Pro Pro Gly Asn
                645                 650                 655
Leu Arg Asn Ser Leu Lys Arg Asp Ala Gln Thr Ile Lys Thr Ile His
            660                 665                 670
Gln Gln Arg Val Leu Pro Ile Glu Gln Ser Leu Ser Thr Leu Tyr Gln
        675                 680                 685
```

```
Ser Val Lys Ile Leu Gln Arg Thr Gly Asn Gly Leu Leu Glu Arg Val
    690                 695                 700
Thr Arg Ile Leu Ala Ser Leu Asp Phe Ala Gln Asn Phe Ile Thr Asn
705                 710                 715                 720
Asn Thr Ser Ser Val Ile Ile Glu Glu Thr Lys Lys Tyr Gly Arg Thr
                725                 730                 735
Ile Ile Gly Tyr Phe Glu His Tyr Leu Gln Trp Ile Glu Phe Ser Ile
            740                 745                 750
Ser Glu Lys Val Ala Ser Cys Lys Pro Val Ala Thr Ala Leu Asp Thr
        755                 760                 765
Ala Val Asp Val Phe Leu Cys Ser Tyr Ile Ile Asp Pro Leu Asn Leu
    770                 775                 780
Phe Trp Phe Gly Ile Gly Lys Ala Thr Val Phe Leu Leu Pro Ala Leu
785                 790                 795                 800
Ile Phe Ala Val Lys Leu Ala Lys Tyr Tyr Arg Arg Met Asp Ser Glu
                805                 810                 815
Asp Val Tyr Asp Asp Val Glu Thr Ile Pro Met Lys Asn Met Glu Asn
            820                 825                 830
Gly Asn Asn Gly Tyr His Lys Asp His Val Tyr Gly Ile His Asn Pro
        835                 840                 845
Val Met Thr Ser Pro Ser Gln His
    850                 855


<210> 3
<211> 856
<212> PRT
<213> Homo sapiens

<400> 3
Met Ala Leu Val Leu Gly Ser Leu Leu Leu Leu Gly Leu Cys Gly Asn
 1               5                  10                  15
Ser Phe Ser Gly Gly Gln Pro Ser Ser Thr Asp Ala Pro Lys Ala Trp
            20                  25                  30
Asn Tyr Glu Leu Pro Ala Thr Asn Tyr Glu Thr Gln Asp Ser His Lys
        35                  40                  45
Ala Gly Pro Ile Gly Ile Leu Phe Glu Leu Val His Ile Phe Leu Tyr
    50                  55                  60
Val Val Gln Pro Arg Asp Phe Pro Glu Asp Thr Leu Arg Lys Phe Leu
65                  70                  75                  80
Gln Lys Ala Tyr Glu Ser Lys Ile Asp Tyr Asp Lys Ile Val Tyr Tyr
                85                  90                  95
Glu Ala Gly Ile Ile Leu Cys Cys Val Leu Gly Leu Leu Phe Ile Ile
            100                 105                 110
Leu Met Pro Leu Val Gly Tyr Phe Phe Cys Met Cys Arg Cys Cys Asn
        115                 120                 125
Lys Cys Gly Gly Glu Met His Gln Arg Gln Lys Glu Asn Gly Pro Phe
    130                 135                 140
Leu Arg Lys Cys Phe Ala Ile Ser Leu Leu Val Ile Cys Ile Ile Ile
145                 150                 155                 160
Ser Ile Gly Ile Phe Tyr Gly Phe Val Ala Asn His Gln Val Arg Thr
                165                 170                 175
Arg Ile Lys Arg Ser Arg Lys Leu Ala Asp Ser Asn Phe Lys Asp Leu
            180                 185                 190
Arg Thr Leu Leu Asn Glu Thr Pro Glu Gln Ile Lys Tyr Ile Leu Ala
        195                 200                 205
Gln Tyr Asn Thr Ile Lys Asp Lys Ala Phe Thr Asp Leu Asn Ser Ile
```

```
    210                 215                 220
Asn Ser Val Leu Gly Gly Gly Ile Leu Asp Arg Leu Arg Pro Asn Ile
225                 230                 235                 240
Ile Pro Val Leu Asp Glu Ile Lys Ser Met Ala Thr Ala Ile Lys Glu
                245                 250                 255
Thr Lys Glu Ala Leu Glu Asn Met Asn Ser Thr Leu Lys Ser Leu His
            260                 265                 270
Gln Gln Ser Thr Gln Leu Ser Ser Ser Leu Thr Ser Val Lys Thr Ser
        275                 280                 285
Leu Arg Ser Ser Leu Asn Asp Pro Leu Cys Leu Val His Pro Ser Ser
    290                 295                 300
Glu Thr Cys Asn Ser Ile Arg Leu Ser Leu Ser Gln Leu Asn Ser Asn
305                 310                 315                 320
Pro Glu Leu Arg Gln Leu Pro Pro Val Asp Ala Glu Leu Asp Asn Val
                325                 330                 335
Asn Asn Val Leu Arg Thr Asp Leu Asp Gly Leu Val Gln Gln Gly Tyr
            340                 345                 350
Gln Ser Leu Asn Asp Ile Pro Asp Arg Val Gln Arg Gln Thr Thr Thr
        355                 360                 365
Val Val Ala Gly Ile Lys Arg Val Leu Asn Ser Ile Gly Ser Asp Ile
    370                 375                 380
Asp Asn Val Thr Gln Arg Leu Pro Ile Gln Asp Ile Leu Ser Ala Phe
385                 390                 395                 400
Ser Val Tyr Val Asn Asn Thr Glu Arg Tyr Ile His Arg Asn Leu Pro
                405                 410                 415
Thr Leu Glu Glu Tyr Asp Ser Tyr Trp Trp Leu Gly Gly Leu Val Ile
            420                 425                 430
Cys Ser Leu Leu Thr Leu Ile Val Ile Phe Tyr Tyr Leu Gly Leu Leu
        435                 440                 445
Cys Gly Val Cys Gly Tyr Asp Arg His Ala Thr Pro Thr Thr Arg Gly
    450                 455                 460
Cys Val Ser Asn Thr Gly Gly Val Phe Leu Met Val Gly Val Gly Leu
465                 470                 475                 480
Ser Phe Leu Phe Cys Trp Ile Leu Met Ile Ile Val Val Leu Thr Phe
                485                 490                 495
Val Phe Gly Ala Asn Val Glu Lys Leu Ile Cys Glu Pro Tyr Thr Ser
            500                 505                 510
Lys Glu Leu Phe Arg Val Leu Asp Thr Pro Tyr Leu Leu Asn Glu Asp
        515                 520                 525
Trp Glu Tyr Tyr Leu Ser Gly Lys Leu Phe Asn Lys Ser Lys Met Lys
    530                 535                 540
Leu Thr Phe Glu Gln Val Tyr Ser Asp Cys Lys Lys Asn Arg Gly Thr
545                 550                 555                 560
Tyr Gly Thr Leu His Leu Gln Asn Ser Phe Asn Ile Ser Glu His Leu
                565                 570                 575
Asn Ile Asn Glu His Thr Gly Ser Ile Ser Ser Glu Leu Glu Ser Leu
            580                 585                 590
Lys Val Asn Leu Asn Ile Phe Leu Leu Gly Ala Ala Gly Arg Lys Asn
        595                 600                 605
Leu Gln Asp Phe Ala Ala Cys Gly Ile Asp Arg Met Asn Tyr Asp Ser
    610                 615                 620
Tyr Leu Ala Gln Thr Gly Lys Ser Pro Ala Gly Val Asn Leu Leu Ser
625                 630                 635                 640
Phe Ala Tyr Asp Leu Glu Ala Lys Ala Asn Ser Leu Pro Pro Gly Asn
                645                 650                 655
Leu Arg Asn Ser Leu Lys Arg Asp Ala Gln Thr Ile Lys Thr Ile His
            660                 665                 670
```

Gln Gln Arg Val Leu Pro Ile Glu Gln Ser Leu Ser Thr Leu Tyr Gln
675 680 685
Ser Val Lys Ile Leu Gln Arg Thr Gly Asn Gly Leu Leu Glu Arg Val
690 695 700
Thr Arg Ile Leu Ala Ser Leu Asp Phe Ala Gln Asn Phe Ile Thr Asn
705 710 715 720
Asn Thr Ser Ser Val Ile Ile Glu Glu Thr Lys Lys Tyr Gly Arg Thr
725 730 735
Ile Ile Gly Tyr Phe Glu His Tyr Leu Gln Trp Ile Glu Phe Ser Ile
740 745 750
Ser Glu Lys Val Ala Ser Cys Lys Pro Val Ala Thr Ala Leu Asp Thr
755 760 765
Ala Val Asp Val Phe Leu Cys Ser Tyr Ile Ile Asp Pro Leu Asn Leu
770 775 780
Phe Trp Phe Gly Ile Gly Lys Ala Thr Val Phe Leu Leu Pro Ala Leu
785 790 795 800
Ile Phe Ala Val Lys Leu Ala Lys Tyr Tyr Arg Arg Met Asp Ser Glu
805 810 815
Asp Val Tyr Asp Asp Val Glu Thr Ile Pro Met Lys Asn Met Glu Asn
820 825 830
Gly Asn Asn Gly Tyr His Lys Asp His Val Tyr Gly Ile His Asn Pro
835 840 845
Val Met Thr Ser Pro Ser Gln His
850 855

<210> 4
<211> 856
<212> PRT
<213> Homo sapiens

<400> 4
Met Ala Leu Val Leu Gly Ser Leu Leu Leu Leu Gly Leu Cys Gly Asn
1 5 10 15
Ser Phe Ser Gly Gly Gln Pro Ser Ser Thr Asp Ala Pro Lys Ala Trp
20 25 30
Asn Tyr Glu Leu Pro Ala Thr Asn Tyr Glu Thr Gln Asp Ser His Lys
35 40 45
Ala Gly Pro Ile Gly Ile Leu Phe Glu Leu Val His Ile Phe Leu Tyr
50 55 60
Val Val Gln Pro Arg Asp Phe Pro Glu Asp Thr Leu Arg Lys Phe Leu
65 70 75 80
Gln Lys Ala Tyr Glu Ser Lys Ile Asp Tyr Asp Lys Ile Val Tyr Tyr
85 90 95
Glu Ala Gly Ile Ile Leu Cys Cys Val Leu Gly Leu Leu Phe Ile Ile
100 105 110
Leu Met Pro Leu Val Gly Tyr Phe Phe Cys Met Cys Arg Cys Cys Asn
115 120 125
Lys Cys Gly Gly Glu Met His Gln Arg Gln Lys Glu Asn Gly Pro Phe
130 135 140
Leu Arg Lys Cys Phe Ala Ile Ser Leu Leu Val Ile Cys Ile Ile Ile
145 150 155 160
Ser Ile Gly Ile Phe Tyr Gly Phe Val Ala Asn His Gln Val Arg Thr
165 170 175
Arg Ile Lys Arg Ser Arg Lys Leu Ala Asp Ser Asn Phe Lys Asp Leu
180 185 190
Arg Thr Leu Leu Asn Glu Thr Pro Glu Gln Ile Lys Tyr Ile Leu Ala

```
            195                     200                     205
Gln Tyr Asn Thr Ile Lys Asp Lys Ala Phe Thr Asp Leu Asn Ser Ile
    210                     215                     220
Asn Ser Val Leu Gly Gly Gly Ile Leu Asp Arg Leu Arg Pro Asn Ile
225                     230                     235                     240
Ile Pro Val Leu Asp Glu Ile Lys Ser Met Ala Thr Ala Ile Lys Glu
                245                     250                     255
Thr Lys Glu Ala Leu Glu Asn Met Asn Ser Thr Leu Lys Ser Leu His
            260                     265                     270
Gln Gln Ser Thr Gln Leu Ser Ser Ser Leu Thr Ser Val Lys Thr Ser
        275                     280                     285
Leu Arg Ser Ser Leu Asn Asp Pro Leu Cys Leu Val His Pro Ser Ser
    290                     295                     300
Glu Thr Cys Asn Ser Ile Arg Leu Ser Leu Ser Gln Leu Asn Ser Asn
305                     310                     315                     320
Pro Glu Leu Arg Gln Leu Pro Pro Val Asp Ala Glu Leu Asp Asn Val
                325                     330                     335
Asn Asn Val Leu Arg Thr Asp Leu Asp Gly Leu Val Gln Gln Gly Tyr
            340                     345                     350
Gln Ser Leu Asn Asp Ile Pro Asp Arg Val Gln Arg Gln Thr Thr Thr
        355                     360                     365
Val Val Ala Gly Ile Lys Arg Val Leu Asn Ser Ile Gly Ser Asp Ile
    370                     375                     380
Asp Asn Val Thr Gln Arg Leu Pro Ile Gln Asp Ile Leu Ser Ala Phe
385                     390                     395                     400
Ser Val Tyr Val Asn Asn Thr Glu Ser Tyr Ile His Arg Asn Leu Pro
                405                     410                     415
Thr Leu Glu Glu Tyr Asp Ser Tyr Trp Trp Leu Gly Gly Leu Val Ile
            420                     425                     430
Cys Ser Leu Leu Thr Leu Ile Val Ile Phe Tyr Tyr Leu Gly Leu Leu
        435                     440                     445
Cys Gly Val Cys Gly Tyr Asp Arg His Ala Thr Pro Thr Thr Arg Gly
    450                     455                     460
Cys Val Ser Asn Thr Gly Gly Val Phe Leu Met Val Gly Val Gly Leu
465                     470                     475                     480
Ser Phe Leu Phe Cys Trp Ile Leu Met Ile Ile Val Val Leu Thr Phe
                485                     490                     495
Val Phe Gly Ala Asn Val Glu Lys Leu Ile Cys Glu Pro Tyr Thr Ser
            500                     505                     510
Lys Glu Leu Phe Arg Val Leu Asp Thr Pro Tyr Leu Leu Asn Glu Asp
        515                     520                     525
Trp Glu Tyr Tyr Leu Ser Gly Lys Leu Phe Asn Lys Ser Lys Met Lys
    530                     535                     540
Leu Thr Phe Glu Gln Val Tyr Ser Asp Cys Lys Lys Asn Arg Gly Thr
545                     550                     555                     560
Tyr Gly Thr Leu His Leu Gln Asn Ser Phe Asn Ile Ser Glu His Leu
                565                     570                     575
Asn Ile Asn Glu His Thr Gly Ser Ile Ser Ser Glu Leu Glu Ser Leu
            580                     585                     590
Lys Val Asn Leu Asn Ile Phe Leu Leu Gly Ala Ala Gly Arg Lys Asn
        595                     600                     605
Leu Gln Asp Phe Ala Ala Cys Gly Ile Asp Arg Met Asn Tyr Asp Ser
    610                     615                     620
Tyr Leu Ala Gln Thr Gly Lys Ser Pro Ala Gly Val Asn Leu Leu Ser
625                     630                     635                     640
Phe Ala Tyr Asp Leu Glu Ala Lys Ala Asn Ser Leu Pro Pro Gly Asn
                645                     650                     655
```

Leu Arg Asn Ser Leu Lys Arg Asp Ala Gln Thr Ile Lys Thr Ile His
660 665 670
Gln Gln Arg Val Leu Pro Ile Glu Gln Ser Leu Ser Thr Leu Tyr Gln
675 680 685
Ser Val Lys Ile Leu Gln Arg Thr Gly Asn Gly Leu Leu Glu Arg Val
690 695 700
Thr Arg Ile Leu Ala Ser Leu Asp Phe Ala Gln Asn Phe Ile Thr Asn
705 710 715 720
Asn Thr Ser Ser Val Ile Ile Glu Glu Thr Lys Lys Tyr Gly Arg Thr
725 730 735
Ile Ile Gly Tyr Phe Glu His Tyr Leu Gln Trp Ile Glu Phe Ser Ile
740 745 750
Ser Glu Lys Val Ala Ser Cys Lys Pro Val Ala Thr Ala Leu Asp Thr
755 760 765
Ala Val Asp Val Phe Leu Cys Ser Tyr Ile Ile Asp Pro Leu Asn Leu
770 775 780
Phe Trp Phe Gly Ile Gly Lys Ala Thr Val Phe Leu Leu Pro Ala Leu
785 790 795 800
Ile Phe Ala Val Lys Leu Ala Lys Tyr Tyr Arg Arg Met Asp Ser Glu
805 810 815
Asp Val Tyr Asp Asp Val Glu Thr Ile Pro Met Lys Asn Met Glu Asn
820 825 830
Gly Asn Asn Gly Tyr His Lys Asp His Val Tyr Gly Ile His Asn Pro
835 840 845
Val Met Thr Ser Pro Ser Gln His
850 855

<210> 5
<211> 856
<212> PRT
<213> Homo sapiens

<400> 5
Met Ala Leu Val Leu Gly Ser Leu Leu Leu Leu Gly Leu Cys Gly Asn
1 5 10 15
Ser Phe Ser Gly Gly Gln Pro Ser Ser Thr Asp Ala Pro Lys Ala Trp
20 25 30
Asn Tyr Glu Leu Pro Ala Thr Asn Tyr Glu Thr Gln Asp Ser His Lys
35 40 45
Ala Gly Pro Ile Gly Ile Leu Phe Glu Leu Val His Ile Phe Leu Tyr
50 55 60
Val Val Gln Pro Arg Asp Phe Pro Glu Asp Thr Leu Arg Lys Phe Leu
65 70 75 80
Gln Lys Ala Tyr Glu Ser Lys Ile Asp Tyr Asp Lys Ile Val Tyr Tyr
85 90 95
Glu Ala Gly Ile Ile Leu Cys Cys Val Leu Gly Leu Leu Phe Ile Ile
100 105 110
Leu Met Pro Leu Val Gly Tyr Phe Phe Cys Met Cys Arg Cys Cys Asn
115 120 125
Lys Cys Gly Gly Glu Met His Gln Arg Gln Lys Glu Asn Gly Pro Phe
130 135 140
Leu Arg Lys Cys Phe Ala Ile Ser Leu Leu Val Ile Cys Ile Ile Ile
145 150 155 160
Ser Ile Gly Ile Phe Tyr Gly Phe Val Ala Asn His Gln Val Arg Thr
165 170 175
Arg Ile Lys Arg Ser Arg Lys Leu Ala Asp Ser Asn Phe Lys Asp Leu

```
                180                     185                     190
Arg Thr Leu Leu Asn Glu Thr Pro Glu Gln Ile Lys Tyr Ile Leu Ala
            195                     200                     205
Gln Tyr Asn Thr Thr Lys Asp Lys Ala Phe Thr Asp Leu Asn Ser Ile
    210                     215                     220
Asn Ser Val Leu Gly Gly Gly Ile Leu Asp Arg Leu Arg Pro Asn Ile
225                     230                     235                     240
Ile Pro Val Leu Asp Glu Ile Lys Ser Met Ala Thr Ala Ile Lys Glu
                245                     250                     255
Thr Lys Glu Ala Leu Glu Asn Met Asn Ser Thr Leu Lys Ser Leu His
            260                     265                     270
Gln Gln Ser Thr Gln Leu Ser Ser Leu Thr Ser Val Lys Thr Ser
        275                     280                     285
Leu Arg Ser Ser Leu Asn Asp Pro Leu Cys Leu Val His Pro Ser Ser
    290                     295                     300
Glu Thr Cys Asn Ser Ile Arg Leu Ser Leu Ser Gln Leu Asn Ser Asn
305                     310                     315                     320
Pro Glu Leu Arg Gln Leu Pro Pro Val Asp Ala Glu Leu Asp Asn Val
                325                     330                     335
Asn Asn Val Leu Arg Thr Asp Leu Asp Gly Leu Val Gln Gln Gly Tyr
            340                     345                     350
Gln Ser Leu Asn Asp Ile Pro Asp Arg Val Gln Arg Gln Thr Thr Thr
        355                     360                     365
Val Val Ala Gly Ile Lys Arg Val Leu Asn Ser Ile Gly Ser Asp Ile
    370                     375                     380
Asp Asn Val Thr Gln Arg Leu Pro Ile Gln Asp Ile Leu Ser Ala Phe
385                     390                     395                     400
Ser Val Tyr Val Asn Asn Thr Glu Arg Tyr Ile His Arg Asn Leu Pro
                405                     410                     415
Thr Leu Glu Glu Tyr Asp Ser Tyr Trp Trp Leu Gly Gly Leu Val Ile
            420                     425                     430
Cys Ser Leu Leu Thr Leu Ile Val Ile Phe Tyr Tyr Leu Gly Leu Leu
        435                     440                     445
Cys Gly Val Cys Gly Tyr Asp Arg His Ala Thr Pro Thr Thr Arg Gly
    450                     455                     460
Cys Val Ser Asn Thr Gly Gly Val Phe Leu Met Val Gly Val Gly Leu
465                     470                     475                     480
Ser Phe Leu Phe Cys Trp Ile Leu Met Ile Ile Val Val Leu Thr Phe
                485                     490                     495
Val Phe Gly Ala Asn Val Glu Lys Leu Ile Cys Glu Pro Tyr Thr Ser
            500                     505                     510
Lys Glu Leu Phe Arg Val Leu Asp Thr Pro Tyr Leu Leu Asn Glu Asp
        515                     520                     525
Trp Glu Tyr Tyr Leu Ser Gly Lys Leu Phe Asn Lys Ser Lys Met Lys
    530                     535                     540
Leu Thr Phe Glu Gln Val Tyr Ser Asp Cys Lys Lys Asn Arg Gly Thr
545                     550                     555                     560
Tyr Gly Thr Leu His Leu Gln Asn Ser Phe Asn Ile Ser Glu His Leu
                565                     570                     575
Asn Ile Asn Glu His Thr Gly Ser Ile Ser Ser Glu Leu Glu Ser Leu
            580                     585                     590
Lys Val Asn Leu Asn Ile Phe Leu Leu Gly Ala Ala Gly Arg Lys Asn
        595                     600                     605
Leu Gln Asp Phe Ala Ala Cys Gly Ile Asp Arg Met Asn Tyr Asp Ser
    610                     615                     620
Tyr Leu Ala Gln Thr Gly Lys Ser Pro Ala Gly Val Asn Leu Leu Ser
625                     630                     635                     640
```

```
Phe Ala Tyr Asp Leu Glu Ala Lys Ala Asn Ser Leu Pro Pro Gly Asn
                645                 650                 655
Leu Arg Asn Ser Leu Lys Arg Asp Ala Gln Thr Ile Lys Thr Ile His
            660                 665                 670
Gln Gln Arg Val Leu Pro Ile Glu Gln Ser Leu Ser Thr Leu Tyr Gln
        675                 680                 685
Ser Val Lys Ile Leu Gln Arg Thr Gly Asn Gly Leu Leu Glu Arg Val
    690                 695                 700
Thr Arg Ile Leu Ala Ser Leu Asp Phe Ala Gln Asn Phe Ile Thr Asn
705                 710                 715                 720
Asn Thr Ser Ser Val Ile Ile Glu Glu Thr Lys Lys Tyr Gly Arg Thr
                725                 730                 735
Ile Ile Gly Tyr Phe Glu His Tyr Leu Gln Trp Ile Glu Phe Ser Ile
            740                 745                 750
Ser Glu Lys Val Ala Ser Cys Lys Pro Val Ala Thr Ala Leu Asp Thr
        755                 760                 765
Ala Val Asp Val Phe Leu Cys Ser Tyr Ile Ile Asp Pro Leu Asn Leu
    770                 775                 780
Phe Trp Phe Gly Ile Gly Lys Ala Thr Val Phe Leu Leu Pro Ala Leu
785                 790                 795                 800
Ile Phe Ala Val Lys Leu Ala Lys Tyr Tyr Arg Arg Met Asp Ser Glu
                805                 810                 815
Asp Val Tyr Asp Asp Val Glu Thr Ile Pro Met Lys Asn Met Glu Asn
            820                 825                 830
Gly Asn Asn Gly Tyr His Lys Asp His Val Tyr Gly Ile His Asn Pro
        835                 840                 845
Val Met Thr Ser Pro Ser Gln His
    850                 855
```

```
<210> 6
<211> 4321
<212> DNA
<213> Homo sapiens

<400> 6
ttttgttgtt gttgtttttt taattgcatt gggattaggc aacagaaggg tctaatgcgg 60
ccgggatgag acaggagagt ttttaggagg gtagctgcgt tctaagtaag ggactctgct 120
ggggtaaaag aggcgcaagc gttgcaagaa gggagtgcag gggttgagc aggcacctct 180
acaggaaatg gatgctgtcc aggtgctggt gggcgcccca gggctacgtg gcgaagcagc 240
tcagccggtc caatcagagt gcgtccaggg ctcgggtttc gcgatcttta agtgactgag 300
gcagatcccc acgcggcacc tggccatgct ctcagctctc ccgccgcggg atggtgcctt 360
gagtgaatga ccccccttgga gaacattctt ccgcatccct cgcctcaagc cagcctcaga 420
cagaaaactg aagattcagc agatccagtg cttcctgctc ctcttctgcc caggaacacg 480
cttgccttcc ccaaggcttc cagaagctct gaggcaggag gcaccaagtt ctacctcatg 540
tttggaggat cttgctagct atggccctcg tactcggctc cctgttgctg ctggggctgt 600
gcgggaactc cttttcagga gggcagcctt catccacaga tgctcctaag gcttggaatt 660
atgaattgcc tgcaacaaat tatgagaccc aagactccca taaagctgga cccattggca 720
ttctctttga actagtgcat atctttctct atgtggtaca gccgcgtgat ttcccagaag 780
atactttgag aaaattctta cagaaggcat atgaatccaa aattgattat gacaagccag 840
aaactgtaat cttaggtcta aagattgtct actatgaagc agggattatt ctatgctgtg 900
tcctggggct gctgtttatt attctgatgc ctctggtggg gtatttcttt tgtatgtgtc 960
gttgctgtaa caaatgtggt ggagaaatgc accagcgaca gaaggaaaat gggcccttcc 1020
tgaggaaatg ctttgcaatc tccctgttgg tgatttgtat aataataagc attggcatct 1080
tctatggttt tgtggcaaat caccaggtaa gaacccggat caaaaggagt cggaaactgg 1140
cagatagcaa tttcaaggac ttgcgaactc tcttgaatga aactccagag caaatcaaat 1200
atatattggc ccagtacaac actaccaagg acaaggcgtt cacagatctg aacagtatca 1260
```

attcagtgct aggaggcgga attcttgacc gactgagacc caacatcatc cctgttcttg 1320
atgagattaa gtccatggca acagcgatca aggagaccaa agaggcgttg gagaacatga 1380
acagcacctt gaagagcttg caccaacaaa gtacacagct tagcagcagt ctgaccagcg 1440
tgaaaactag cctgcggtca tctctcaatg accctctgtg cttggtgcat ccatcaagtg 1500
aaacctgcaa cagcatcaga ttgtctctaa gccagctgaa tagcaaccct gaactgaggc 1560
agcttccacc cgtggatgca gaacttgaca acgttaataa cgttcttagg acagatttgg 1620
atggcctggt ccaacagggc tatcaatccc ttaatgatat acctgacaga gtacaacgcc 1680
aaaccacgac tgtcgtagca ggtatcaaaa gggtcttgaa ttccattggt tcagatatcg 1740
acaatgtaac tcagcgtctt cctattcagg atatactctc agcattctct gtttatgtta 1800
ataacactga aagttacatc cacagaaatt tacctacatt ggaagagtat gattcatact 1860
ggtggctggg tggcctggtc atctgctctc tgctgaccct catcgtgatt ttttactacc 1920
tgggcttact gtgtggcgtg tgcggctatg acaggcatgc caccccgacc acccgaggct 1980
gtgtctccaa caccggaggc gtcttcctca tggttggagt tggattaagt ttcctctttt 2040
gctggatatt gatgatcatt gtggttctta cctttgtctt tggtgcaaat gtggaaaaac 2100
tgatctgtga accttacacg agcaaggaat tattccgggt tttggataca ccctacttac 2160
taaatgaaga ctgggaatac tatctctctg ggaagctatt taataaatca aaaatgaagc 2220
tcacttttga acaagtttac agtgactgca aaaaaaatag aggcacttac ggcactcttc 2280
acctgcagaa cagcttcaat atcagtgaac atctcaacat taatgagcat actggaagca 2340
taagcagtga attggaaagt ctgaaggtaa atcttaatat ctttctgttg ggtgcagcag 2400
gaagaaaaaa ccttcaggat tttgctgctt gtggaataga cagaatgaat tatgacagct 2460
acttggctca gactggtaaa tcccccgcag gagtgaatct tttatcattt gcatatgatc 2520
tagaagcaaa agcaaacagt ttgcccccag gaaatttgag gaactccctg aaaagagatg 2580
cacaaactat taaaacaatt caccagcaac gagtccttcc tatagaacaa tcactgagca 2640
ctctatacca aagcgtcaag atacttcaac gcacagggaa tggattgttg gagagagtaa 2700
ctaggattct agcttctctg gattttgctc agaacttcat cacaaacaat acttcctctg 2760
ttattattga ggaaactaag aagtatggga gaacaataat aggatatttt gaacattatc 2820
tgcagtggat cgagttctct atcagtgaga aagtggcatc gtgcaaacct gtggccaccg 2880
ctctagatac tgctgttgat gtctttctgt gtagctacat tatcgacccc ttgaatttgt 2940
tttggtttgg cataggaaaa gctactgtat ttttacttcc ggctctaatt tttgcggtaa 3000
aactggctaa gtactatcgt cgaatggatt cggaggacgt gtacgatgat gttgaaacta 3060
tacccatgaa aaatatggaa aatggtaata atggttatca taaagatcat gtatatggta 3120
ttcacaatcc tgttatgaca agcccatcac aacattgata gctgatgttg aaactgcttg 3180
agcatcagga tactcaaagt ggaaaggatc acagattttt ggtagtttct gggtctacaa 3240
ggactttcca aatccaggag caacgccagt ggcaacgtag tgactcaggc gggcaccaag 3300
gcaacggcac cattggtctc tgggtagtgc tttaagaatg aacacaatca cgttatagtc 3360
catggtccat cactattcaa ggatgactcc ctcccttcct gtctattttt gtttttttact 3420
tttttacact gagtttctat ttagacacta caacatatgg ggtgtttgtt cccattggat 3480
gcatttctat caaaactcta tcaaatgtga tggctagatt ctaacatatt gccatgtgtg 3540
gagtgtgctg aacacacacc agtttacagg aaagatgcat tttgtgtaca gtaaacggtg 3600
tatatacctt ttgttaccac agagtttttt aaacaaatga gtattatagg actttcttct 3660
aaatgagcta aataagtcac cattgacttc ttggtgctgt tgaaaataat ccattttcac 3720
taaaagtgtg tgaaacctac agcatattct tcacgcagag attttcatct attatacttt 3780
atcaaagatt ggccatgttc cacttggaaa tggcatgcaa aagcaatcat agagaaacct 3840
gcgtaactcc atctgacaaa ttcaaaagag agagagagat cttgagagag aaatgctgtt 3900
cgttcaaaag tggagttgtt ttaacagatg ccaattacgg tgtacagttt aacagagttt 3960
tctgttgcat taggataaac attaattgga gtgcagctaa catgagtatc atcagactag 4020
tatcaagtgt tctaaaatga aatatgagaa gatcctgtca caattcttag atctggtgtc 4080
cagcatggat gaaacctttg agtttggtcc ctaaatttgc atgaaagcac aaggtaaata 4140
ttcatttgct tcaggagttt catgttggat ctgtcattat caaaagtgat cagcaatgaa 4200
gaactggtcg gacaaaattt aacgttgatg taatggaatt ccagatgtag gcattccccc 4260
caggtctttt catgtgcaga ttgcagttct gattcatttg aataaaaagg aacttggaaa 4320
a 4321

<210> 7
<211> 2571
<212> DNA
<213> Homo sapiens

<400> 7
atggccctcg tactcggctc cctgttgctg ctggggctgt gcgggaactc cttttcagga 60
gggcagcctt catccacaga tgctcctaag gcttggaatt atgaattgcc tgcaacaaat 120
tatgagaccc aagactccca taaagctgga cccattggca ttctctttga actagtgcat 180
atctttctct atgtggtaca gccgcgtgat ttcccagaag atactttgag aaaattctta 240
cagaaggcat atgaatccaa aattgattat gacaagattg tctactatga agcagggatt 300
attctatgct gtgtcctggg gctgctgttt attattctga tgcctctggt ggggtatttc 360
ttttgtatgt gtcgttgctg taacaaatgt ggtggagaaa tgcaccagcg acagaaggaa 420
aatgggccct tcctgaggaa atgctttgca atctccctgt tggtgatttg tataataata 480
agcattggca tcttctatgg ttttgtggca aatcaccagg taagaacccg gatcaaaagg 540
agtcggaaac tggcagatag caatttcaag gacttgcgaa ctctcttgaa tgaaactcca 600
gagcaaatca aatatatatt ggcccagtac aacactacca aggacaaggc gttcacagat 660
ctgaacagta tcaattcagt gctaggaggc ggaattcttg accgactgag acccaacatc 720
atccctgttc ttgatgagat taagtccatg gcaacagcga tcaaggagac caaagaggcg 780
ttggagaaca tgaacagcac cttgaagagc ttgcaccaac aaagtacaca gcttagcagc 840
agtctgacca gcgtgaaaac tagcctgcgg tcatctctca atgaccctct gtgcttggtg 900
catccatcaa gtgaaacctg caacagcatc agattgtctc taagccagct gaatagcaac 960
cctgaactga ggcagcttcc acccgtggat gcagaacttg acaacgttaa taacgttctt 1020
aggacagatt tggatggcct ggtccaacag ggctatcaat cccttaatga tatacctgac 1080
agagtacaac gccaaaccac gactgtcgta gcaggtatca aaagggtctt gaattccatt 1140
ggttcagata tcgacaatgt aactcagcgt cttcctattc aggatatact ctcagcattc 1200
tctgtttatg ttaataacac tgaaagttac atccacagaa atttacctac attggaagag 1260
tatgattcat actggtggct gggtggcctg gtcatctgct ctctgctgac cctcatcgtg 1320
atttttttact acctgggctt actgtgtggc gtgtgcggct atgacaggca tgccaccccg 1380
accacccgag gctgtgtctc caacaccgga ggcgtcttcc tcatggttgg agttggatta 1440
agtttcctct tttgctggat attgatgatc attgtggttc ttacctttgt ctttggtgca 1500
aatgtggaaa aactgatctg tgaaccttac acgagcaagg aattattccg ggttttggat 1560
acaccctact tactaaatga agactgggaa tactatctct ctgggaagct atttaataaa 1620
tcaaaaatga agctcacttt tgaacaagtt tacagtgact gcaaaaaaaa tagaggcact 1680
tacggcactc ttcacctgca gaacagcttc aatatcagtg aacatctcaa cattaatgag 1740
catactggaa gcataagcag tgaattggaa agtctgaagg taaatcttaa tatctttctg 1800
ttgggtgcag caggaagaaa aaaccttcag gattttgctg cttgtggaat agacagaatg 1860
aattatgaca gctacttggc tcagactggt aaatcccccg caggagtgaa tcttttatca 1920
tttgcatatg atctagaagc aaaagcaaac agtttgcccc caggaaattt gaggaactcc 1980
ctgaaaagag atgcacaaac tattaaaaca attcaccagc aacgagtcct tcctatagaa 2040
caatcactga gcactctata ccaaagcgtc aagatacttc aacgcacagg gaatggattg 2100
ttggagagag taactaggat tctagcttct ctggattttg ctcagaactt catcacaaac 2160
aatacttcct ctgttattat tgaggaaact aagaagtatg ggagaacaat aataggatat 2220
tttgaacatt atctgcagtg gatcgagttc tctatcagtg agaaagtggc atcgtgcaaa 2280
cctgtggcca ccgctctaga tactgctgtt gatgtctttc tgtgtagcta cattatcgac 2340
cccttgaatt tgttttggtt tggcatagga aaagctactg tatttttact tccggctcta 2400
atttttgcgg taaaactggc taagtactat cgtcgaatgg attcggagga cgtgtacgat 2460
gatgttgaaa ctatacccat gaaaaatatg gaaaatggta ataatggtta tcataaagat 2520
catgtatatg gtattcacaa tcctgttatg acaagcccat cacaacatta g 2571

<210> 8
<211> 2571
<212> DNA
<213> Homo sapiens

<400> 8
atggccctcg tactcggctc cctgttgctg ctggggctgt gcgggaactc cttttcagga 60
gggcagcctt catccacaga tgctcctaag gcttggaatt atgaattgcc tgcaacaaat 120
tatgagaccc aagactccca taaagctgga cccattggca ttctctttga actagtgcat 180
atctttctct atgtggtaca gccgcgtgat ttcccagaag atactttgag aaaattctta 240
cagaaggcat atgaatccaa aattgattat gacaagattg tctactatga agcagggatt 300

```
attctatgct gtgtcctggg gctgctgttt attattctga tgcctctggt ggggtatttc 360
ttttgtatgt gtcgttgctg taacaaatgt ggtggagaaa tgcaccagcg acagaaggaa 420
aatgggccct tcctgaggaa atgctttgca atctccctgt tggtgatttg tataataata 480
agcattggca tcttctatgg ttttgtggca aatcaccagg taagaacccg gatcaaaagg 540
agtcggaaac tggcagatag caatttcaag gacttgcgaa ctctcttgaa tgaaactcca 600
gagcaaatca aatatatatt ggcccagtac aacactatca aggacaaggc gttcacagat 660
ctgaacagta tcaattcagt gctaggaggc ggaattcttg accgactgag acccaacatc 720
atccctgttc ttgatgagat taagtccatg gcaacagcga tcaaggagac caaagaggcg 780
ttggagaaca tgaacagcac cttgaagagc ttgcaccaac aaagtacaca gcttagcagc 840
agtctgacca gcgtgaaaac tagcctgcgg tcatctctca atgaccctct gtgcttggtg 900
catccatcaa gtgaaacctg caacagcatc agattgtctc taagccagct gaatagcaac 960
cctgaactga ggcagcttcc acccgtggat gcagaacttg acaacgttaa taacgttctt 1020
aggacagatt tggatggcct ggtccaacag ggctatcaat cccttaatga tatacctgac 1080
agagtacaac gccaaaccac gactgtcgta gcaggtatca aaagggtctt gaattccatt 1140
ggttcagata tcgacaatgt aactcagcgt cttcctattc aggatatact ctcagcattc 1200
tctgtttatg ttaataacac tgaaaggtac atccacagaa atttacctac attggaagag 1260
tatgattcat actggtggct gggtggcctg gtcatctgct ctctgctgac cctcatcgtg 1320
atttttatact acctgggctt actgtgtggc gtgtgcggct atgacaggca tgccaccccg 1380
accacccgag gctgtgtctc caacaccgga ggcgtcttcc tcatggttgg agttggatta 1440
agtttcctct tttgctggat attgatgatc attgtggttc ttacctttgt ctttggtgca 1500
aatgtggaaa aactgatctg tgaaccttac acgagcaagg aattattccg ggttttggat 1560
acaccctact tactaaatga agactgggaa tactatctct ctgggaagct atttaataaa 1620
tcaaaaatga agctcacttt tgaacaagtt tacagtgact gcaaaaaaaa tagaggcact 1680
tacggcactc ttcacctgca gaacagcttc aatatcagtg aacatctcaa cattaatgag 1740
catactggaa gcataagcag tgaattggaa agtctgaagg taaatcttaa tatctttctg 1800
ttgggtgcag caggaagaaa aaaccttcag gattttgctg cttgtggaat agacagaatg 1860
aattatgaca gctacttggc tcagactggt aaatccccccg caggagtgaa tcttttatca 1920
tttgcatatg atctagaagc aaaagcaaac agtttgcccc caggaaattt gaggaactcc 1980
ctgaaaagag atgcacaaac tattaaaaca attcaccagc aacgagtcct tcctatagaa 2040
caatcactga gcactctata ccaaagcgtc aagatacttc aacgcacagg gaatggattg 2100
ttggagagag taactaggat tctagcttct ctggattttg ctcagaactt catcacaaac 2160
aatacttcct ctgttattat tgaggaaact aagaagtatg ggagaacaat aataggatat 2220
tttgaacatt atctgcagtg gatcgagttc tctatcagtg agaaagtggc atcgtgcaaa 2280
cctgtggcca ccgctctaga tactgctgtt gatgtctttc tgtgtagcta cattatcgac 2340
cccttgaatt tgttttggtt tggcatagga aaagctactg tatttttact tccggctcta 2400
atttttgcgg taaaactggc taagtactat cgtcgaatgg attcggagga cgtgtacgat 2460
gatgttgaaa ctatacccat gaaaaatatg gaaaatggta ataatggtta tcataaagat 2520
catgtatatg gtattcacaa tcctgttatg acaagcccat cacaacatta g          2571
```

<210> 9
<211> 2571
<212> DNA
<213> Homo sapiens

<400> 9

```
atggccctcg tactcggctc cctgttgctg ctggggctgt gcgggaactc cttttcagga 60
gggcagcctt catccacaga tgctcctaag gcttggaatt atgaattgcc tgcaacaaat 120
tatgagaccc aagactccca taaagctgga cccattggca ttctctttga actagtgcat 180
atctttctct atgtggtaca gccgcgtgat ttcccagaag atactttgag aaaattctta 240
cagaaggcat atgaatccaa aattgattat gacaagattg tctactatga agcagggatt 300
attctatgct gtgtcctggg gctgctgttt attattctga tgcctctggt ggggtatttc 360
ttttgtatgt gtcgttgctg taacaaatgt ggtggagaaa tgcaccagcg acagaaggaa 420
aatgggccct tcctgaggaa atgctttgca atctccctgt tggtgatttg tataataata 480
agcattggca tcttctatgg ttttgtggca aatcaccagg taagaacccg gatcaaaagg 540
agtcggaaac tggcagatag caatttcaag gacttgcgaa ctctcttgaa tgaaactcca 600
gagcaaatca aatatatatt ggcccagtac aacactatca aggacaaggc gttcacagat 660
ctgaacagta tcaattcagt gctaggaggc ggaattcttg accgactgag acccaacatc 720
```

atccctgttc ttgatgagat taagtccatg gcaacagcga tcaaggagac caaagaggcg 780
ttggagaaca tgaacagcac cttgaagagc ttgcaccaac aaagtacaca gcttagcagc 840
agtctgacca gcgtgaaaac tagcctgcgg tcatctctca atgaccctct gtgcttggtg 900
catccatcaa gtgaaacctg caacagcatc agattgtctc taagccagct gaatagcaac 960
cctgaactga ggcagcttcc acccgtggat gcagaacttg acaacgttaa taacgttctt 1020
aggacagatt tggatggcct ggtccaacag ggctatcaat cccttaatga tatacctgac 1080
agagtacaac gccaaaccac gactgtcgta gcaggtatca aaagggtctt gaattccatt 1140
ggttcagata tcgacaatgt aactcagcgt cttcctattc aggatatact ctcagcattc 1200
tctgtttatg ttaataacac tgaaagttac atccacagaa atttacctac attggaagag 1260
tatgattcat actggtggct gggtggcctg gtcatctgct ctctgctgac cctcatcgtg 1320
attttttact acctgggctt actgtgtggc gtgtgcggct atgacaggca tgccaccccg 1380
accacccgag gctgtgtctc caacaccgga ggcgtcttcc tcatggttgg agttggatta 1440
agtttcctct tttgctggat attgatgatc attgtggttc ttacctttgt ctttggtgca 1500
aatgtggaaa aactgatctg tgaaccttac acgagcaagg aattattccg ggttttggat 1560
acaccctact tactaaatga agactgggaa tactatctct ctgggaagct atttaataaa 1620
tcaaaaatga agctcacttt tgaacaagtt tacagtgact gcaaaaaaaa tagaggcact 1680
tacggcactc ttcacctgca gaacagcttc aatatcagtg aacatctcaa cattaatgag 1740
catactggaa gcataagcag tgaattggaa agtctgaagg taaatcttaa tatctttctg 1800
ttgggtgcag caggaagaaa aaaccttcag gattttgctg cttgtggaat agacagaatg 1860
aattatgaca gctacttggc tcagactggt aaatcccccg caggagtgaa tcttttatca 1920
tttgcatatg atctagaagc aaaagcaaac agtttgcccc caggaaattt gaggaactcc 1980
ctgaaaagag atgcacaaac tattaaaaca attcaccagc aacgagtcct tcctatagaa 2040
caatcactga gcactctata ccaaagcgtc aagatacttc aacgcacagg gaatggattg 2100
ttggagagag taactaggat tctagcttct ctggattttg ctcagaactt catcacaaac 2160
aatacttcct ctgttattat tgaggaaact aagaagtatg ggagaacaat aataggatat 2220
tttgaacatt atctgcagtg gatcgagttc tctatcagtg agaaagtggc atcgtgcaaa 2280
cctgtggcca ccgctctaga tactgctgtt gatgtctttc tgtgtagcta cattatcgac 2340
cccttgaatt tgttttggtt tggcatagga aaagctactg tatttttact tccggctcta 2400
atttttgcgg taaaactggc taagtactat cgtcgaatgg attcggagga cgtgtacgat 2460
gatgttgaaa ctatacccat gaaaaatatg gaaaatggta ataatggtta tcataaagat 2520
catgtatatg gtattcacaa tcctgttatg acaagcccat cacaacatta g 2571

<210> 10
<211> 2571
<212> DNA
<213> Homo sapiens

<400> 10
atggccctcg tactcggctc cctgttgctg ctggggctgt gcgggaactc cttttcagga 60
gggcagcctt catccacaga tgctcctaag gcttggaatt atgaattgcc tgcaacaaat 120
tatgagaccc aagactccca taaagctgga cccattggca ttctctttga actagtgcat 180
atctttctct atgtggtaca gccgcgtgat ttcccagaag atactttgag aaaattctta 240
cagaaggcat atgaatccaa aattgattat gacaagattg tctactatga agcagggatt 300
attctatgct gtgtcctggg gctgctgttt attattctga tgcctctggt ggggtatttc 360
ttttgtatgt gtcgttgctg taacaaatgt ggtggagaaa tgcaccagcg acagaaggaa 420
aatgggccct tcctgaggaa atgctttgca atctccctgt tggtgatttg tataataata 480
agcattggca tcttctatgg ttttgtggca aatcaccagg taagaacccg gatcaaaagg 540
agtcggaaac tggcagatag caatttcaag gacttgcgaa ctctcttgaa tgaaactcca 600
gagcaaatca aatatatatt ggcccagtac aacactacca aggacaaggc gttcacagat 660
ctgaacagta tcaattcagt gctaggaggc ggaattcttg accgactgag acccaacatc 720
atccctgttc ttgatgagat taagtccatg gcaacagcga tcaaggagac caaagaggcg 780
ttggagaaca tgaacagcac cttgaagagc ttgcaccaac aaagtacaca gcttagcagc 840
agtctgacca gcgtgaaaac tagcctgcgg tcatctctca atgaccctct gtgcttggtg 900
catccatcaa gtgaaacctg caacagcatc agattgtctc taagccagct gaatagcaac 960
cctgaactga ggcagcttcc acccgtggat gcagaacttg acaacgttaa taacgttctt 1020
aggacagatt tggatggcct ggtccaacag ggctatcaat cccttaatga tatacctgac 1080
agagtacaac gccaaaccac gactgtcgta gcaggtatca aaagggtctt gaattccatt 1140

ggttcagata tcgacaatgt aactcagcgt cttcctattc aggatatact ctcagcattc 1200
tctgtttatg ttaataacac tgaaaggtac atccacagaa atttacctac attggaagag 1260
tatgattcat actggtggct gggtggcctg gtcatctgct ctctgctgac cctcatcgtg 1320
atttttttact acctgggctt actgtgtggc gtgtgcggct atgacaggca tgccaccccg 1380
accaccccgag gctgtgtctc caacaccgga ggcgtcttcc tcatggttgg agttggatta 1440
agtttcctct tttgctggat attgatgatc attgtggttc ttacctttgt ctttggtgca 1500
aatgtggaaa aactgatctg tgaaccttac acgagcaagg aattattccg ggttttggat 1560
acaccctact tactaaatga agactgggaa tactatctct ctgggaagct atttaataaa 1620
tcaaaaatga agctcacttt tgaacaagtt tacagtgact gcaaaaaaaa tagaggcact 1680
tacggcactc ttcacctgca gaacagcttc aatatcagtg aacatctcaa cattaatgag 1740
catactggaa gcataagcag tgaattggaa agtctgaagg taaatcttaa tatctttctg 1800
ttgggtgcag caggaagaaa aaaccttcag gattttgctg cttgtggaat agacagaatg 1860
aattatgaca gctacttggc tcagactggt aaatcccccg caggagtgaa tcttttatca 1920
tttgcatatg atctagaagc aaaagcaaac agtttgcccc caggaaattt gaggaactcc 1980
ctgaaaagag atgcacaaac tattaaaaca attcaccagc aacgagtcct tcctatagaa 2040
caatcactga gcactctata ccaaagcgtc aagatacttc aacgcacagg gaatggattg 2100
ttggagagag taactaggat tctagcttct ctggattttg ctcagaactt catcacaaac 2160
aatacttcct ctgttattat tgaggaaact aagaagtatg ggagaacaat aataggatat 2220
tttgaacatt atctgcagtg gatcgagttc tctatcagtg agaaagtggc atcgtgcaaa 2280
cctgtggcca ccgctctaga tactgctgtt gatgtctttc tgtgtagcta cattatcgac 2340
cccttgaatt tgttttggtt tggcatagga aaagctactg tatttttact tccggctcta 2400
atttttgcgg taaaactggc taagtactat cgtcgaatgg attcggagga cgtgtacgat 2460
gatgttgaaa ctatacccat gaaaaatatg gaaaatggta ataatggtta tcataaagat 2520
catgtatatg gtattcacaa tcctgttatg acaagcccat cacaacatta g 2571

<210> 11
<211> 3794
<212> RNA
<213> Homo sapiens

<400> 11
ccaagttcta cctcatgttt ggaggatctt gctagctatg gccctcgtac tcggctccct 60
gttgctgctg gggctgtgcg ggaactcctt ttcaggaggg cagccttcat ccacagatgc 120
tcctaaggct tggaattatg aattgcctgc aacaaattat gagacccaag actcccataa 180
agctggaccc attggcattc tctttgaact agtgcatatc tttctctatg tggtacagcc 240
gcgtgatttc ccagaagata ctttgagaaa attcttacag aaggcatatg aatccaaaat 300
tgattatgac aagccagaaa ctgtaatctt aggtctaaag attgtctact atgaagcagg 360
gattattcta tgctgtgtcc tggggctgct gtttattatt ctgatgcctc tggtggggta 420
tttcttttgt atgtgtcgtt gctgtaacaa atgtggtgga gaaatgcacc agcgacagaa 480
ggaaaatggg cccttcctga ggaaatgctt tgcaatctcc ctgttggtga tttgtataat 540
aataagcatt ggcatcttct atggttttgt ggcaaatcac caggtaagaa cccggatcaa 600
aaggagtcgg aaactggcag atagcaattt caaggacttg cgaactctct tgaatgaaac 660
tccagagcaa atcaaatata tattggccca gtacaacact accaaggaca aggcgttcac 720
agatctgaac agtatcaatt cagtgctagg aggcggaatt cttgaccgac tgagacccaa 780
catcatccct gttcttgatg agattaagtc catggcaaca gcgatcaagg agaccaaaga 840
ggcgttggag aacatgaaca gcaccttgaa gagcttgcac caacaaagta cacagcttag 900
cagcagtctg accagcgtga aaactagcct gcggtcatct ctcaatgacc ctctgtgctt 960
ggtgcatcca tcaagtgaaa cctgcaacag catcagattg tctctaagcc agctgaatag 1020
caaccctgaa ctgaggcagc ttccacccgt ggatgcagaa cttgacaacg ttaataacgt 1080
tcttaggaca gatttggatg gcctggtcca acagggctat caatccctta atgatatacc 1140
tgacagagta caacgccaaa ccacgactgt cgtagcaggt atcaaaaggg tcttgaattc 1200
cattggttca gatatcgaca atgtaactca gcgtcttcct attcaggata tactctcagc 1260
attctctgtt tatgttaata acactgaaag ttacatccac agaaatttac ctacattgga 1320
agagtatgat tcatactggt ggctgggtgg cctggtcatc tgctctctgc tgaccctcat 1380
cgtgattttt tactacctgg gcttactgtg tggcgtgtgc ggctatgaca ggcatgccac 1440
cccgaccacc cgaggctgtg tctccaacac cggaggcgtc ttcctcatgg ttggagttgg 1500
attaagtttc ctcttttgct ggatattgat gatcattgtg gttcttacct ttgtctttgg 1560

```
tgcaaatgtg gaaaaactga tctgtgaacc ttacacgagc aaggaattat tccgggtttt 1620
ggatacaccc tacttactaa atgaagactg ggaatactat ctctctggga agctatttaa 1680
taaatcaaaa atgaagctca cttttgaaca agtttacagt gactgcaaaa aaaatagagg 1740
cacttacggc actcttcacc tgcagaacag cttcaatatc agtgaacatc tcaacattaa 1800
tgagcatact ggaagcataa gcagtgaatt ggaaagtctg aaggtaaatc ttaatatctt 1860
tctgttgggt gcagcaggaa gaaaaaacct tcaggatttt gctgcttgtg gaatagacag 1920
aatgaattat gacagctact tggctcagac tggtaaatcc cccgcaggag tgaatctttt 1980
atcatttgca tatgatctag aagcaaaagc aaacagtttg cccccaggaa atttgaggaa 2040
ctccctgaaa agagatgcac aaactattaa aacaattcac cagcaacgag tccttcctat 2100
agaacaatca ctgagcactc tataccaaag cgtcaagata cttcaacgca cagggaatgg 2160
attgttggag agagtaacta ggattctagc ttctctggat tttgctcaga acttcatcac 2220
aaacaatact tcctctgtta ttattgagga aactaagaag tatgggagaa caataatagg 2280
atattttgaa cattatctgc agtggatcga gttctctatc agtgagaaag tggcatcgtg 2340
caaacctgtg gccaccgctc tagatactgc tgttgatgtc tttctgtgta gctacattat 2400
cgaccccttg aatttgtttt ggtttggcat aggaaaagct actgtatttt tacttccggc 2460
tctaattttt gcggtaaaac tggctaagta ctatcgtcga atggattcgg aggacgtgta 2520
cgatgatgtt gaaactatac ccatgaaaaa tatggaaaat ggtaataatg gttatcataa 2580
agatcatgta tatggtattc acaatcctgt tatgacaagc ccatcacaac attgatagct 2640
gatgttgaaa ctgcttgagc atcaggatac tcaaagtgga aaggatcaca gatttttggt 2700
agtttctggg tctacaagga ctttccaaat ccaggagcaa cgccagtggc aacgtagtga 2760
ctcaggcggg caccaaggca acggcaccat tggtctctgg gtagtgcttt aagaatgaac 2820
acaatcacgt tatagtccat ggtccatcac tattcaagga tgactccctc ccttcctgtc 2880
tatttttgtt ttttactttt ttacactgag tttctattta gacactacaa catatggggt 2940
gtttgttccc attggatgca tttctatcaa aactctatca aatgtgatgg ctagattcta 3000
acatattgcc atgtgtggag tgtgctgaac acacaccagt ttacaggaaa gatgcatttt 3060
gtgtacagta aacggtgtat ataccttttg ttaccacaga gtttttttaaa caaatgagta 3120
ttataggact ttcttctaaa tgagctaaat aagtcaccat tgacttcttg gtgctgttga 3180
aaataatcca ttttcactaa aagtgtgtga aacctacagc atattcttca cgcagagatt 3240
ttcatctatt atactttatc aaagattggc catgttccac ttggaaatgg catgcaaaag 3300
ccatcataga gaaacctgcg taactccatc tgacaaattc aaaagagaga gagagatctt 3360
gagagagaaa tgctgttcgt tcaaaagtgg agttgtttta acagatgcca attacggtgt 3420
acagtttaac agagttttct gttgcattag gataaacatt aattggagtg cagctaacat 3480
gagtatcatc agactagtat caagtgttct aaaatgaaat atgagaagat cctgtcacaa 3540
ttcttagatc tggtgtccag catggatgaa acctttgagt ttggtcccta aatttgcatg 3600
aaagcacaag gtaaatattc atttgcttca ggagtttcat gttggatctg tcattatcaa 3660
aagtgatcag caatgaagaa ctggtcggac aaaatttaac gttgatgtaa tggaattcca 3720
gatgtaggca ttcccccag gtcttttcat gtgcagattg cagttctgat tcatttgaat 3780
aaaaaggaac ttgg                                                   3794
```